

Deep Learning in Whole-Slide Imaging for Precision Computer-Aided Diagnosis

Journal:	<i>IEEE Journal of Biomedical and Health Informatics</i>
Manuscript ID	JBHI-01482-2019
Manuscript Type:	Special Issue on Computational Pathology
Date Submitted by the Author:	30-Nov-2019
Complete List of Authors:	Tong, Li; Georgia Institute of Technology, Biomedical Engineering Choudhary, Anirudh; Georgia Institute of Technology, Computational Science and Engineering Malakarjun Patil, Shreyas; Georgia Institute of Technology, Electrical and Computer Engineering Zhang, Yundong Zhu, Yuanda; Georgia Tech, School of Electrical and Computer Engineering; Georgia Tech, The Wallace H. Coulter Department of Biomedical Engineering Wang, May; Georgia Institute of Technology, Biomedical Eng.

SCHOLARONE™
Manuscripts

Deep Learning in Whole-Slide Imaging for Precision Computer-Aided Diagnosis

Li Tong, Anirudh Choudhary[†], Shreyas Malakarjun Patil[†], Yundong Zhang[†], Yuanda Zhu, and May D. Wang, *Senior Member, IEEE*

Abstract—Pathology plays a key role in modern medicine for disease diagnosis. However, traditional pathology reading is subjective with inconsistency. With recent advances in microscopic imaging and advanced AI, computational digital pathology (i.e., an entire tissue biopsy is scanned in as a whole-slide image, a.k.a. WSI, followed by quantitative analysis done by computer) has emerged and grown rapidly. Because each WSI contains a huge number of pixels, AI techniques such as Deep Learning (DL) have been utilized to accomplish data-driven WSI segmentation and classification. In this article, we critique how state-of-the-art DL algorithms have revolutionized computational pathology. First, we review the state-of-art WSI analysis using DL, including preprocessing, data quality control, and various DL model training (e.g., supervised, semi-supervised, and weakly supervised learning). Second, we discuss advanced DL topics for WSI such as model interpretation and multi-modal data integration. Last, for grand challenges in applying DL to WSI such as the lack of labeled data, the heterogeneity of whole-slide images, and the lack of model transparency, we summarize major opportunities for both biomedical engineering and pathology communities to further advance computational pathology.

Index Terms—Whole slide imaging, pathological images, deep learning, computer aided diagnosis, precision medicine

I. INTRODUCTION

PRECISION medicine and personalized treatment have drawn increasing expectations in health care, especially in cancer treatment. Pathologists are central to this public expectation, as they are the first to make decisions regarding the patients. Recent studies, nevertheless, pointed out two challenges concerning pathologists. Bainbridge et al. [1] showed that 44% of pathologists work overtime weekly in the UK; meanwhile, active physicians in pathology decreased by 10.4% between 2008 and 2013, and over 60% of active

[†]These authors contributed equally to this work.

L. Tong and MD Wang are with the Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, 30332. E-mail (MD Wang): maywang@gatech.edu

A. Choudhary is with the Department of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, 30332

S. Malakarjun Patil and Y. Zhu are with the School of Electronic Engineering, Georgia Institute of Technology, Atlanta, GA, 30332

The work was supported in part by grants from the National Center for Advancing Translational Sciences of the National Institute of Health (NIH) under Award UL1TR000454, the National Science Foundation EAGER Award NSF1651360, Children's Healthcare of Atlanta and Georgia Tech Partnership Grant, Giglio Breast Cancer Research Fund, the Centers for Disease Control and Prevention (CDC), and Carol Ann and David D. Flanagan Faculty Fellow Research Fund. This work was also supported in part by the scholarship from China Scholarship Council (CSC) under the Grant CSC NO. 201406010343. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Manuscript received April 19, 2005; revised August 26, 2015.

pathologists are 55 years or older [2]. Computational digital pathology (**Figure 1A**) is the solution as it streamlines the workflow in screening tasks and even provides a computer-aided diagnosis (CAD) system as a reference, significantly reducing the workload of pathologists.

Whole slide imaging (WSI) scans a complete glass slide into a multi-scale high-resolution digital file. The adoption of whole slide images has enabled easy storage, management, sharing, and analysis of the pathological images. Thus, WSI has been widely utilized in computational pathology for automated and quantitative computer-aided diagnosis. In conventional digital image processing, the general pipeline for WSI analysis consists of quality control, feature extraction, predictive modeling, and clinical decision making [3], [4]. Among these steps, the feature extraction step is most challenging and primarily determined the overall performance of the pipeline. With conventional digital imaging processing techniques, hand-crafted pathological imaging features are constructed at pixel-level and object-level, respectively. Although these features can be further filtered by feature selection methods in a data-driven fashion, the design of these features heavily relies on domain knowledge. Thus, the representation power of conventional digital imaging processing techniques is limited, which significantly restricts the performance and generalization capacity of the pipeline.

With the availability of big data and the exponential increase in the computational capability (especially GPUs), deep learning (DL) [5], characterized by nonlinear deep neural networks such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has achieved superior performance in various tasks, including computer vision [6] and natural language processing [7], [8], in comparison with the conventional approaches. By learning sophisticated representations from a large amount of data, DL works with higher dimensional spaces and is a purely data-driven approach. The unreasonable effectiveness is also being portrayed when applied to biomedical data including medical images (e.g., MRI [9], CT [10], pathological images [11]), electronic health records [12], and multi-omics profiles [13].

The trend has also been observed in WSI analysis. The application of DL in digital pathology has revolutionized the field and achieved state-of-the-art performances. Previous challenges in shallow learners such as inter-dataset variance, super-high resolution, and time-complexities are being solved with DL. As a natural extension from computer vision, CNNs have been utilized in computational pathology for diseases classification, nuclei and cell segmentation, and cell counting.

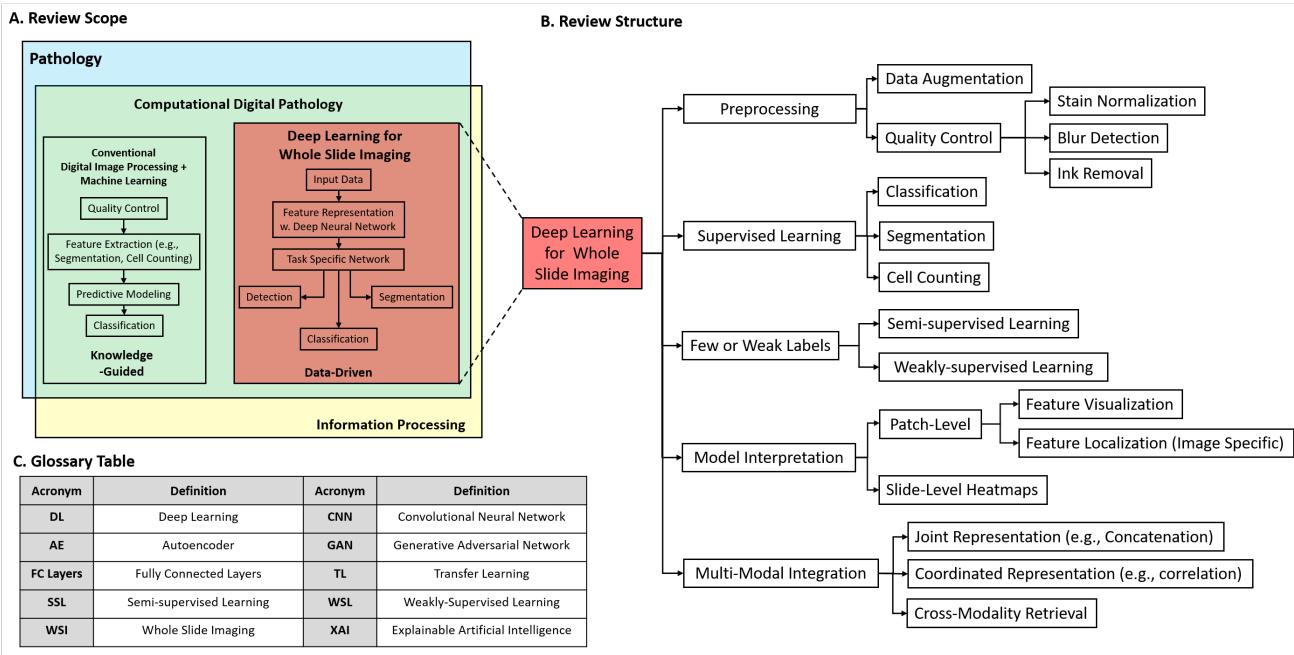


Fig. 1. The review scope and review Structure of this article. A. Review scope. As part of the computational digital pathology, our review will focus on deep learning for whole slide imaging. B. Review structure. C. Glossary Table.

Some well-known CNN architectures such as ResNet, U-Net, and convolutional autoencoders have achieved excellent performance in these tasks. As shown in **Figure 2** with an exemplary pipeline for whole slide classification, the DL pipeline for whole slide images consists of preprocessing, patch-level classification, and slide-level classification. For preprocessing, we first sample image patches from the slide and then apply quality control and data augmentation. For patch-level classification, CNNs and fully connected layers are usually applied together for feature representation and image classification. For slide-level classification, various feature aggregation approaches are utilized to combine the patch-level information for making a combined decision.

However, the performance of DL models heavily relies on the amount of well-annotated WSI data. Also, more advanced DL architectures are needed to address the input size constraints and to improve the slide-level predictions. On the other hand, the field of DL in WSI just took off in recent years, and hence many territories remain unexplored, such as deep reinforcement learning, RNNs, and generative adversarial networks (GANs) based architectures. For example, while the use of GANs is quite prevalent in computer vision and natural images, it has failed to gain momentum in pathological images except for preprocessing.

Recent studies [14], [15] have reported some challenges and opportunities for applying DL to pathology. However, DL methodologies for WSI have not been systematically reviewed before. In this paper, we are going to review the DL approaches developed or applied for WSI informatics (**Figure 1**). In addition, we will perform a comprehensive review on potential opportunities and constraints when applying DL specifically on WSIs. These DL approaches are critical in computational pathology, providing higher coverage,

extracting better features and making more accurate decisions regarding whole-slide images. Ultimately they serve better health care by promoting precision medicine and personalized care.

This paper is organized as follows. In **Section II**, we will review the DL approaches for whole-slide images preprocessing and quality control. We will then discuss the supervised, unsupervised, semi-supervised, and weakly-supervised DL approaches for WSI in **Sections III** and **IV**, respectively. In **Section V** and **Section VI**, we will discuss the model interpretation and multi-modal integration for WSIs, respectively. Lastly, we will summarize the challenges and opportunities for DL-based WSI analysis in **Section VII**.

II. PREPROCESSING AND QUALITY CONTROL

A. Data Augmentation

Developing a robust and generalizable supervised approach using deep learning requires access to a large amount of annotated training data covering various possible input variations. However, manual generation of ground truth training data for tasks such as nuclei segmentation and region-of-interest annotation is a tedious task. Thus, most pathology datasets comprise of limited number of samples. Researchers rely on synthetic data augmentation to introduce artificial variations using color and morphological transformations. These variations account for the potential deviations in future test data making the model robust. Stain augmentation focuses on generating synthetic images by modifying color-related properties or perturbing the stain vectors [16] [17]. Mahmood et al. [18] generated a large labeled dataset for nuclei segmentation using unpaired cycle-consistent adversarial network to transform randomly generated polygonal segmentation masks

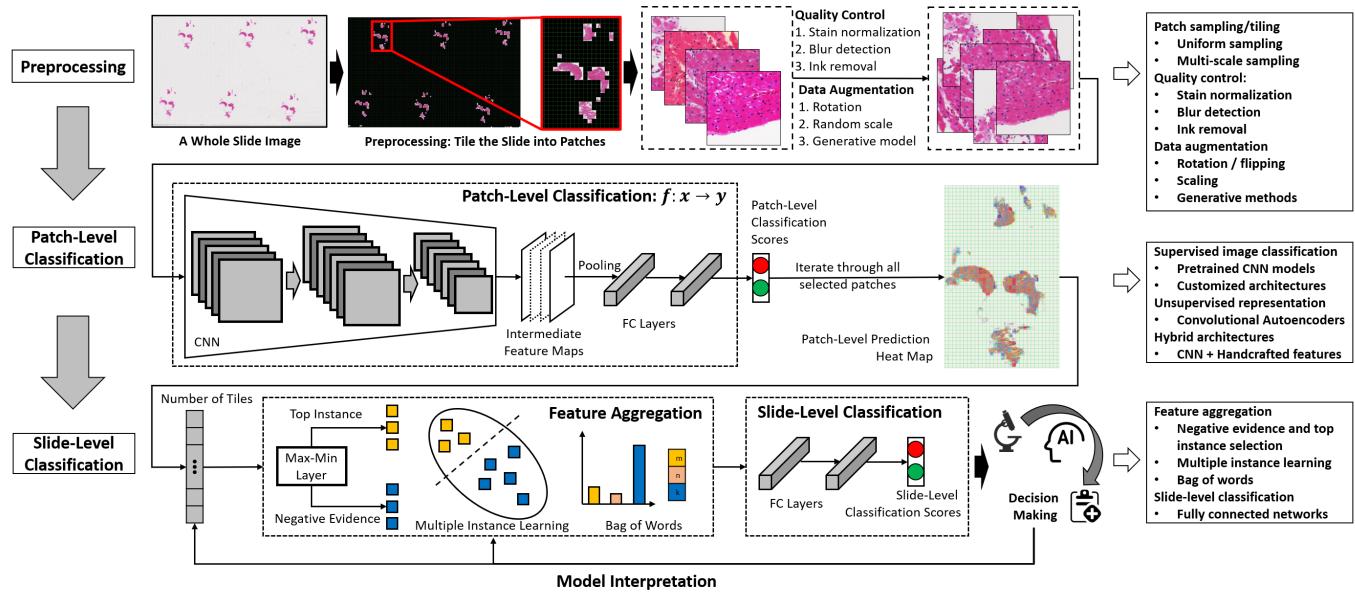


Fig. 2. Exemplary pipeline of DL-based classification for whole-slide images. The main steps include pre-processing, patch-level classification, and slide-level classification

into realistic WSI patches. Hou et al. [19] generated synthetic training patches for nuclei segmentation by utilizing tissue texture information from real patches and refining artificial images using GAN [20]. They highlight two novel approaches for effective downstream analysis 1) using importance sampling-based weights for synthetic images during training, 2) forcing the generator to create hard-to-classify synthetic samples. [21] use data augmentation for stain-invariant segmentation by generating virtual image samples for each stain domain except that of the input image domain and concatenating the input image with virtual images. They leverage CycleGAN to learn the domain-to-domain stain mapping.

B. Quality Control

Histopathology whole slide image generation involves multiple processes which can introduce artifacts and lead to heterogeneity among images acquired at different sites, known as batch-effect. Both these issues can induce bias and have unpredictable effects on downstream medical applications like image segmentation and classification [22]. With the advent of collaborative data repositories involving high throughput data from multiple institutions, quality control becomes an essential step in the whole-slide image processing pipeline for developing robust computer-aided diagnosis (CAD) systems. In this section, we review DL-based pre-processing techniques addressing two major areas: batch-effect normalization and artifact removal.

1) Batch-Effect Normalization: One of the most common batch-effects in WSI is stain variation. Stain variations are caused by inconsistencies in the digitization process such as varying stain procedures, different image scanners, and stain compositions. Conventional stain normalization methods rely on pixel-level distribution matching or stain vector estimation to derive the color transformation function. To account for

spatial dependency, researchers try to account for semantic information by learning independent transformations for tissue components [23] or performing joint normalization [24].

Compared to conventional approaches, DL techniques formulate stain normalization as an unsupervised domain translation problem, jointly learning the stain transfer mapping between different batches (Figure 3). DL methods utilize generative models to effectively represent the distribution-shift, leveraging both source and target domains completely. CNNs have been shown to be more effective in capturing semantic and color related perceptual information compared to pixel-based methods [25]. Most stain normalization methods draw upon generative models popular in natural image domain for multi-domain image-to-image translation, image generation and semantic manipulation. Some approaches perform stain color alignment jointly with optimizing for the downstream task (segmentation/classification) formulating it as a supervised learning problem. Moreover, stain normalization can also be considered as a special case of virtual re-staining i.e. artificially generating stained images with multiple chemical stains. Table I provides a summary of various stain normalization methods.

Explicit distribution matching: Explicit distribution matching methods try to parameterize the underlying color distribution by assuming prior distribution family. DL-based stain normalization methods cluster the image into tissue sub-types by learning latent feature representations and perform independent color distribution alignment using estimated posteriors. Color alignment is generally performed using histogram-matching or SVD. Zanjani et. al. [26] trained variational autoencoder (VAE) [27] on image reconstruction to learn color distribution using unit Gaussian prior. They also proposed a likelihood maximization based approach which achieves superior results, using GMM to perform soft clustering on

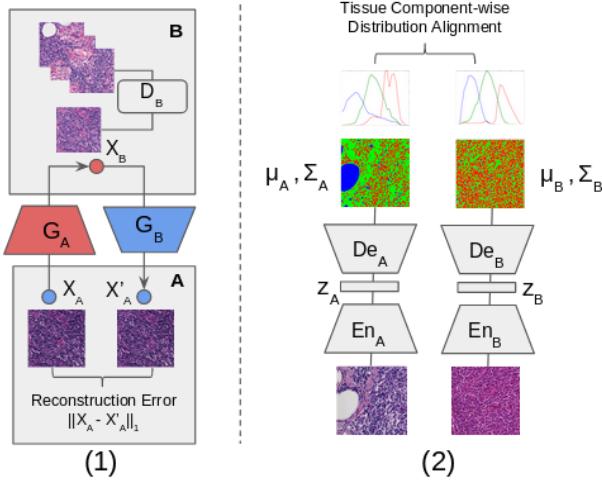


Fig. 3. Overview of two major approaches for stain normalization of whole slide images. (1) Implicit distribution learning using CycleGAN (2) Explicit distribution matching using clustering

CNN features derived from optical density image and aligning multivariate Gaussians in HSD color space using SVD. [28] used unsupervised sparse autoencoder (StaNoSA) [29] to cluster RGB image high-dimensional feature embeddings. While clustering-based approaches don't require tissue semantics related assumptions or manual parameter tuning, their reliance on cluster count might lead to sub-optimal performance in case of imbalanced representation of tissue components.

Implicit distribution learning: Implicit distribution learning methods directly sample from the target data without explicitly assuming a prior parametric distribution. GANs can implicitly learn the color distribution mapping and inspired from natural images, pix2pix and CycleGAN are most widely used architectures for image translation. Cycle-consistent GAN (CycleGAN) [30] are effective for unpaired image translation by enforcing identity loss for semantic consistency and is widely used for unsupervised stain normalization. [31] proposed the first approach using vanilla CycleGAN, named StainGAN. Virtual re-staining studies have incorporated additional constraints into CycleGAN to preserve perceptual and discriminative details such as embedding-consistency loss [32] and photo-realism & structural-similarity loss [33]. [32] developed an overlapping-window based color transfer approach to ensure tile-consistency. Most studies address one-to-one domain scenario, however pathology tissues can be sourced from multiple sites and recent works have focused on learning joint mapping for multiple stain domains. [34] formulate stain normalization as a paired image-colorization problem [35] using encoder-decoder network while [33] customize CycleGAN to map input images to a new stain domain imposing constraints for distinct discriminator boundary. Inspired from pix2pix architecture, [36] and [37] leverage class information to train conditional GAN with U-Net based encoder for paired color translation. Limited work exploring alternative GAN architectures is available, including [38] which incorporate self-attention after convolutional layers in CycleGAN to preserve finer tissue details and [26] who leveraged InfoGAN to learn latent noise

features for tissue components but achieve sub-optimal results. Performing stain normalization independently might incorporate non-relevant variability features which affect downstream performance and augmenting stain transfer with supervised learning has been shown to lead to improved performance [39]. [40] proposed one of the initial approaches leveraging adversarial domain adaptation approach [41] by training GAN for stain normalization imposing an edge-weighted structure preserving loss and class conditioning the generator for cancer classification. They achieve more than 10% improvement in classification performance post normalization and showcase the impact of joint optimization on classifier's generalization ability as well as quality of normalized images. [39] incorporate feature-preserving loss between input and normalized images generated using GAN and showcase higher accuracy on cancer classification. [32] employ a distance-based loss on the latent embedding learned by encoder while [33] enforce tissue category-based feature consistency by optimizing for cancer classification and passing category information to the discriminator. Stain normalization on color augmented data can further improve classification performance as shown by Tellez et. al. [16]. They perform extensive color augmentation in HSV & HED spaces and utilize UNet-based architecture to perform stain transfer on augmented data. Another approach combining task's true performance-based learning with style reconstruction was proposed by [42]. [43] investigated two different approaches for performing stain-independent segmentation, and found that performing stain-normalization to a reference domain on which segmentation model has been trained leads to higher accuracy, particularly if the reference stain enhances tissue component separability.

While multiple loss formulations have been explored to preserve tissue micro-structure, limited work has been done on formulating efficient distance formulations for comparing distributions, such as Earth Movers Distance which has been found to be effective for style [44] and color [45] transfer.

Evaluation Methods: Evaluating color normalization methods is challenging due to perceptual factors involved. Most studies perform qualitative analysis or evaluate impact on downstream task performance post normalization. However, qualitative assessments are typically small-scale and lack validation studies with well-trained pathologists, except in [33]. Quantitative assessment relies on structural metrics like SSIM, CWSSIM, FSSIM to measure image quality, which don't account for varying tissue structure between unpaired images. Alternative approaches compare stain distribution for H&E stained regions using statistical measures like NMI [26] but require manual annotation of tissue regions to generate ground-truth data. While deep networks perform better than many traditional approaches, whether clustering or GAN-based approach is better is still inconclusive due to the lack of standard benchmarking framework and dataset.

2) *Tissue Artifacts:* Errors in biopsy slide preparation or in microscope parameters may lead to tissue artifacts such as tissue folds, blurred regions, pen marks, shadows, and chromatic aberrations. Image artifacts lead to sub-optimal image features [54] and need to be eliminated or rectified. However, very few studies have leveraged DL for resolving

TABLE I
A SUMMARY OF STATE-OF-THE-ART DL-BASED PREPROCESSING METHODS FOR HISTOPATHOLOGY IMAGES

Task	Method	Description	Examples
Stain Normalization			
Style Transfer	Image translation using cycle-consistent GAN	Image translation to reference stain domain with L_1 cycle-consistency constraint on source domain images. Additional losses incorporated to preserve tissue micro-structures such as structural-similarity loss, feature-embedding L_2 loss and reconstruction loss.	[31], [46], [32], [47], [38], [48]
	Image translation using InfoGAN	Modeling the noise vectors for stain normalization	[49]
	Image colorization using Encoder-Decoder framework	Transformation to color-invariant space(grayscale) or shared color space and learning mapping to reference stain space	[34]
Style Transfer + Auxiliary Task	Adversarial image translation with classification	Jointly optimize stain normalization with downstream task ensuring that the discriminative features are retained during normalization	[33], [42], [40], [39]
Explicit Distribution Matching	Tissue clustering using CNN and Gaussian Mixture Model	Learn color-related parameters for tissue components by clustering image using CNN-derived features and likelihood maximization	[26]
	Feature Learning with Autoencoder	Learn latent color features for tissue sub-types using variational autoencoder for image-reconstruction or cluster tissue components using feature encoding learnt using sparse autoencoder	[26], [28]
Tissue Artifact Removal			
Tissue Fold	Convolutional feature learning	Classify tissue fold patches using CNN	[50]
Deblur	Convolutional feature learning	Identify blurred regions using CNN features and kNN-based classifier	[51]
	Image translation using SRGAN	Generate high resolution images	[52]
Ink Removal	ROI Detection (CNN+YOLO)	Segment ink-marked patches and restore them using CycleGAN	[53]

tissue artifacts.

Tissue Fold: Tissue-fold is caused by layering of non-adherent tissue on the slide and has been identified traditionally using color and intensity-based thresholding. However, color variations within WSI can lead to failure of color feature-based approaches. Babaie et. al. [50] trained SVM classifier with CNN-based feature extractor for identifying tissue-fold patches. They highlight that improved data augmentation (varying patch-sizes, patch-types and tissue-fold proportion) could lead to better generalization.

Deblur / Out-of-focus correction: Low resolution tissue images are caused due to blurring or out-of-focus scanner or access to low-resolution devices. Removing blur or recovering high resolution images is non-trivial, due to lack of paired training data. Mukherjee et. al. [51] proposed a CNN-based framework, enhanced by a dictionary-based K-Nearest Neighbor framework during the post-processing of CNN output, to generate high resolution images. Deblurring is also approached as super-resolution image generation task and SRGAN (Super Resolution Generative Adversarial Network) has been a highly effective architecture in Computer Vision [55]. However, SRGAN uses mean squared error (MSE) as the loss function, making the model overly-sensitive to the training data. To adapt the GAN-based framework to histopathology images, Upadhyay et al [52] proposed the quasi-norm based loss function; making it robust to data corruption (blurring) and outperforming baseline SRGAN-based model. Their network is capable of modeling the rich textural features and learning the complicated mappings between low and high resolution images.

Ink Markings Removal: Pathologists routinely highlight certain regions of WSI with permanent ink markings for downstream analysis. These ink markings contaminate the training patches and might deteriorate the performance of DL models. Ali et al [53] proposed a fully automatic pipeline to recover original WSI patches by locating markings using CNN-based binary classifier and YOLO-based detector. They

leverage CycleGAN to translate ink-markers patches to generate realistic original WSI patches.

III. SUPERVISED DEEP LEARNING

DL has known to achieve exceptional results in visual recognition tasks such as classification, object detection, and semantic segmentation. In the grand scheme of things, the availability of a huge number of labeled data plays a crucial role. Thus, DL is framed as a data-driven approach. All the tasks in machine learning can be condensed down to fitting a function and finding the underlying probability distribution through some example data points. The most straightforward setting among these tasks is the supervised learning. Supervised learning refers to fitting a function with respect to data that contains both input and output to the function. For supervised learning in WSI, the task varies with respect to the input of different scales of whole-slide images and the output of different targets (e.g., classification, segmentation, and cell counting) (**Table II**). Due to the enormous size of WSI, the direct application of DL to the entire slide at a high-resolution level is difficult. Hence, the analysis is typically first applied to the image patches sampled from the slide and then to the whole-slide.

A. Classification

Classification tasks in histopathology WSI are concerned with the diagnosis of sub-type cancers. A comparison study between the handcrafted features and deep features learning has been provided in [56]. The study experimentally proves that deep features provide better results than conventional feature extraction methods. In this section, we will first introduce the related DL concepts and then summarize their applications for classification in WSI.

Representations and Deep Neural Networks Embedding features from the input space to representations that are discriminating for the task at hand allows for better classification

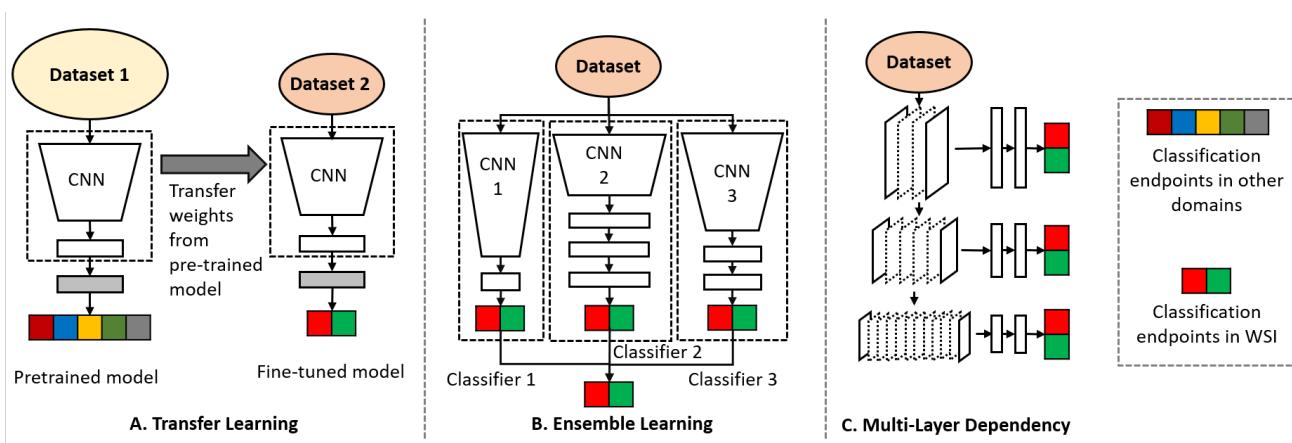


Fig. 4. Overview of the techniques for DL-based classification of whole-slide images. A. Transfer learning: initiate the weights from pre-trained models and fine-tune the parameters with limited number of labeled data. B. Ensemble learning: combine the decisions of several CNN-based classifiers. C. Multi-layer dependencies: utilize the dependencies between consequent layers for classification.

performance. Cruz in [57] and Xu in [58] present a similar approach for classification using structured DNNs. Also, when learning to increase the dissimilarity between various classes in a representation space is applied the accuracy goes up as portrayed in [59]. Works have also shown progress in the adaptation of feature spaces to histology features, in [60] fisher vectors have been adapted to the representations.

Convolutional Neural Networks (CNNs) CNNs have been proved as one of the best classifiers in modern visual recognition problems and are the current state-of-the-art. The obstacles of applying CNNs to WSI are two-fold. First, the input size of high-resolution whole-slide images is enormous, and direct feeding into CNNs is prohibitive. Second, the labels are available on the slide-level and hence the training on patches or samples poses a challenge. [11] overcame this challenge by maximizing the expectation of patches being discriminating and then using this result in an SVM for a slide-level training. CNNs have been applied to classification in histopathology data for various diseases, such as gastric carcinoma [61], lung tumour [62], and colorectal polyps [63]. The state-of-the-art CNN architectures such as VGG [64], GoogleNet [65] and ResNet [66] have been applied for classification [63].

Domain Adaptation Domain adaptation or transfer learning refers to the learning across domains. The most straightforward approach is to initialize the weights learned from other domains (e.g., ImageNet) and fine-tune the weights with respect to current application (**Figure 4A**). Researchers have applied transfer learning and ensemble of CNNs at various resolution levels for the classification of pathological images [17], [67], [68]. Domain adaption is also applied to the augmented data, which has been generated by random sampling and rotations [69]. Motlag et al. develop a hierarchical application of ResNet50 and ResNet 152 [66] pre-trained models, for cancer organ classification and sub-type disease classification respectively [70]. Bayramoglu et al. develop a transfer-learning-based approach for cell nuclei classification using CNN given limited data [17].

Ensemble Learning The classification for WSI is usually a

two-step process, wherein several weak classifiers are trained on a patch-level and then the results are processed together using methods such as voting or gradient boosting. In the era of DL, these classifiers have moved from conventional SVMs to CNNs (**Figure 4B**). Gupta et al. present a joint color-texture feature extraction and a classifier ensemble for classifying breast histopathology images [71], [72]. Similarly the use of gradient boosting tree classifiers is shown in [73]. On the other hand, Vang et al. utilizes an ensemble fusion framework involving majority voting, gradient boosting machine (GBM), and logistic regression exploits. Golatkar et al. [75] uses voting on top of CNNs. The most common CNN seen in these approaches is the InceptionV3 [91].

Multi-Layer Dependencies There exists dependencies among layers in deep neural networks. As the data passes through each layer, the forward layer learns finer features on top of the learned features. Methods have been proposed to exploit these dependencies (**Figure 4C**). Gupta et al. propose a fine-tuned Dense-Net architecture [76], where every layer has the capability of making decisions depending upon whether output score qualifies a predefined confidence threshold. Similarly, Gupta et al. propose to integrate multi-layered features while training ResNet [77] so as to select an optimal sub-set of layers for prediction. Integration between various layers and also various patches from the same WSI can be learned by combining residual layers and recurring layers. Alom et al. developed an inception recurrent residual CNN architecture to achieve the same [78].

B. Segmentation

Segmentation refers to identification of region for a particular artifact and its delineation. The task is to predict a binary mask, hence is depicted as a pixel-wise classification problem. Several deep CNN based approaches have been proposed for semantic segmentation, including Mask RCNN [113] and U-Net [88]. The major challenges for the use of these methods on WSI are the colossal size of the input and the lack of annotated samples. However, researchers have demonstrated successful

TABLE II
SUPERVISED DEEP LEARNING METHODS FOR WSI

Task	Methods	Description	Examples
Classification			
Patch-Level Classification	Representation and Feature engineering using DNNs	Classifiers on top of Features extracted using CNNs	[57] [58] [59] [60]
	Convolutional Neural Networks	Direct applications of various state-of-the-art networks on histology images for classification.	[11] [61] [62] [63]
	Transfer Learning	Fine-tuning of parameters learned over a different domain, usually a very large data-set.	[17] [67] [68] [69] [70] [66] [17]
	Ensemble of several classifiers	Several classifiers are trained from scratch and then the results are processed together using methods such as voting or gradient boosting	[71] [72] [73] [74] [75]
	Integration of Multi-layer dependencies	Dependency exists among the layers in deep networks. Several methods have been proposed, which exploit these dependencies.	[76] [77] [78]
Segmentation			
Cell Segmentation (Patch-Level Segmentation)	Pixel-wise Classification using CNNs	Multi-tasking approach, both classification of the patch as well as segmentation of cell.	[79] [80] [81] [82] [83] [84] [85] [86]
	Pixel-wise classification using Auto-encoders	Downsampling of input and then reconstruction of a binary mask.	[87] [88] [89] [90]
	Generative modelling for binary masks	Conditional GAN has been correspondingly used to manipulate the segmentation tasks for nuclei.	[18]
Metastasis Detection (Slide-Level Segmentation)	ROI proposal using CNN	Sampling of patches and then classifying for metastasis of patches, CNNs used - AlexNet [6], GoogLeNet [65], VGG16 [64], InceptionV3 [91], and ResNetv4 [92]	[93]–[97] [98] [99], [100]
Cell Counting / Object Detection			
Cell Counting	Detection-based	Performs an object detection on the patch-level images and then counts the number of detected cells	[101] [102]
	Regression-based	Maps to a cell density map and estimates the cell count from the density map.	[103] [104] [105] [106] [107]
Mitosis Detection	Feature ensemble	Combining the feature learning from CNN with handcrafted features for efficiency	[108] [105]
	Pixel-wise Classification	CNNs either in the form of an encoder-decoder approach or maintain the same resolution level by using appropriate padding	[109] [110]
	Localization	Network predicts four coordinates with respect to the vertices of the rectangle.	[111] [112]

application of DL for segmentation on WSI at various scales. The segmentation in WSI is divided into two levels, first is the patch-level segmentation, such as cell and nuclei segmentation. The second is the slide-level segmentation, such as the region of interest (ROI) proposal. Previous research has demonstrated that features learned from deep CNNs outperform handcrafted features based approaches in discriminating regions through classification [81].

1) *Patch-Level Segmentation:* The patch-level segmentation aims to discriminate structures at the cellular level, such as segmenting cells or nuclei. Some major challenges are, the large variance in texture, shape, and size of cells make accurate delineation difficult. Second, the cells can be dense and overlapped. Thus, a clear boundary and distinction between two cells is strenuous. Third, image artifacts such as blurring can lead to challenges such as unclear cell boundaries.

Cell Segmentation: The accurate delineation of cells in histopathology is essential for the diagnosis of cancers. Machine learning approaches using CNN based architectures such as autoencoders and GANs have been developed and applied to image patches sampled from the WSI.

- **CNNs:** Segmentation using CNNs on samples transforms to a multi-tasking approach, which does both classification of the patch as well as segmentation of cell inside the patch. Some examples of CNNs being used in such a manner are presented in [79]–[86].

- **Autoencoders:** In [87], a combination of residual network architecture and autoencoders has been proposed for the segmentation task. Ronneberger et al. proposed U-Net, a generalized autoencoder based architecture, for biomedical image segmentation [88]. The U-Net architecture utilizes skip connections between the encoder outputs and decoder at the same resolution level. It has been used several times for cell segmentation [89], [90].

- **GANs:** Generative adversarial networks are very rare in the task of semantic segmentation due to the unpredictable nature and also the higher data requirements than other DL approaches. An approach has been presented that uses synthetic data to augment and meet the dataset size requirements. A conditional GAN has been correspondingly used to manipulate the segmentation tasks for nuclei [18].

2) *Slide-level Segmentation:* The slide-level segmentation tasks in WSI mainly correspond to the region of interest (ROI) proposal. The major challenges for ROI proposal in WSI are the size and scales of WSI, which leads to segmentation on a down-sampled image (i.e., thumbnail) or iteratively on small image patches. Working on the low-resolution thumbnail is similar to natural images, which has been reviewed in patch-level. The sampling of WSI compresses the task to binary classification of the patches as belonging to the ROI. A few semantic segmentation based architectures have been

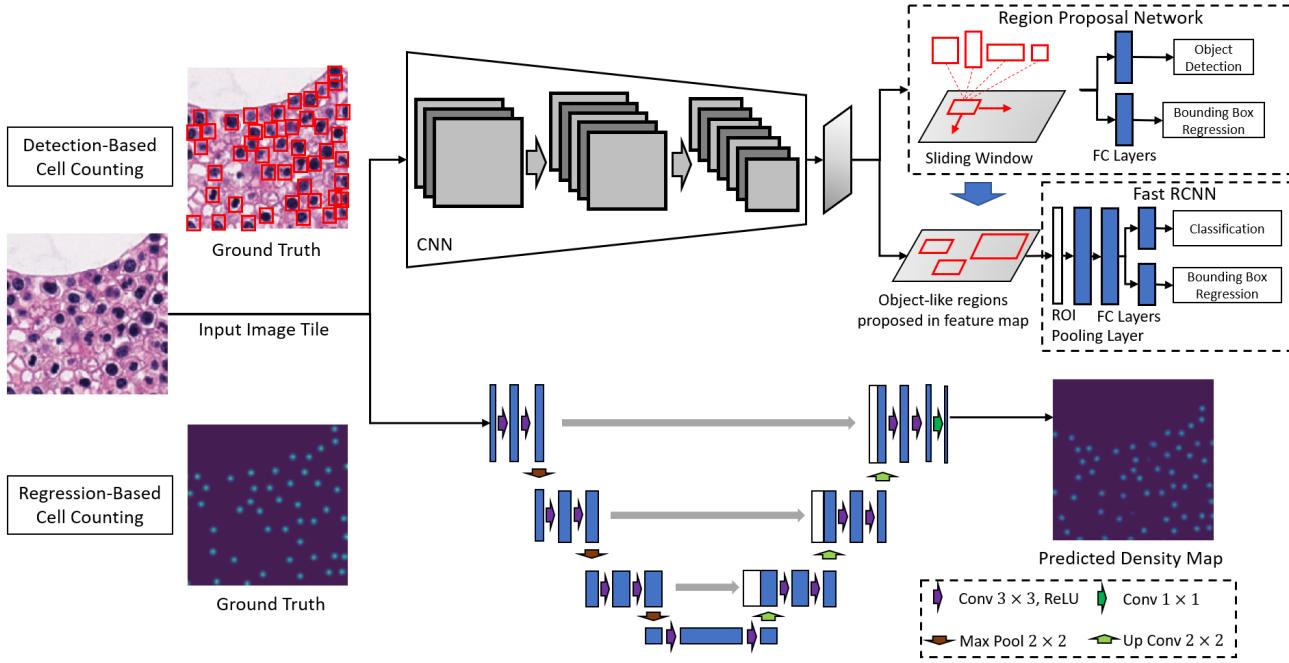


Fig. 5. Detection-based and regression-based cell counting. The detection-based cell counting requires bounding-box annotations and can be implemented with the Faster Region-CNN (R-CNN) architecture. The regression-based cell counting only needs a dot annotation in the center of each cell and usually adopts a U-Net architecture.

applied to slide-level segmentation [114]. In this study, various architectures, include FCN-8s, two SegNet variants, and multi-scale U-Net [88], have been tested. Another such approach is presented in [115]. However, this process can be computationally inefficient and time-consuming. Proper sampling strategies reduce the computational load while maintaining accuracy. For example, Madabhushi et al. developed a high through-put adaptive sampling process for invasive cancer region proposals [116]. On the other hand, pathologists generally assess on a coarser level and then zoom in to regions that are skeptical. Inspired by the procedure, Dong et al. propose a framework that can learn whether zooming-in is required for different regions using CNNs [117].

Metastasis Detection: The methods proposed or metastasis detection follow a similar pipeline [93]–[97]. First, image patches are sampled from the slide. Next, a DL-based classifier is used for identifying the patches with metastasis. Although the basic pipeline is similar, the CNN structures and their depth has evolved significantly. The earlier approaches used AlexNet [6], and then evolved to GoogLeNet [65], VGG16 [64], InceptionV3 [91], and ResNetv4 [92]. The corresponding depth associated with these networks increased chronologically. A more evolved CNN based approach is presented in [98], where the authors first eliminate frequent outliers and then classify the filtered patches [99], [100].

C. Cell Counting

Patch-level cell counting is an important aspect of WSI analysis. Given an input image patch x , we aim to calculate the number of cells $y = f(x)$ with a DL model f . The cell counting task can be challenging because of the various

types, morphological patterns, and distributions of different cells. Several public benchmarks for pathological imaging cell counting use a single pixel to represent each cell with the center of the cell labeled as 1 and 0 otherwise [103]. Because one image patch may contain hundreds of cells, the dot-annotation minimizes the efforts required for labeling. The cell counting in can be achieved in two categories of DL methods: detection and regression (**Figure 5**).

1) *Detection-based cell counting:* The procedure first performs an object detection on the patch-level images and then counts the number of detected cells. For example, Paeng et al. use cell density as a metric for the region of interest proposal and then classify the cells under mitosis with fully convolutional NN [101]. Hung et al. have applied Faster R-CNN for counting [102]. The advantage is the ability to count different type of cells. However, the detection-based cell counting can be challenging because of the occlusion for densely distributed cells and the annotation intensive process.

2) *Regression-based cell counting:* The regression methods maps to a cell density map and estimates the cell count from the density map. In conventional image processing, cell density map can be estimated with handcrafted features. For example, Kainz et al has applied a regression random forest to 21 handcrafted features and predicted for each pixel the distance to the closest cell center. The cell centers can then be identified by finding local maximums and summed counting [103]. DNNs have been applied to square sliding windows as a feature extractor and a classifier for each pixel [104]. Wang et al. further extended this work by combining handcrafted features with CNN features to improve the performance and speed [105]. The sliding-window approach is time-consuming

and limits generalization to cells of various shapes and sizes. The input-output mapping can also be achieved by U-Net [88]. To apply U-Net for regression-based cell counting, we first smooth the single-pixel annotated ground truth with a Gaussian kernel. Then, the cell counting can be achieved by summing the predicted cell density map. For example, Guo et al. has developed a universal cell counting model by improving the U-Net architecture with self-attention and online batch normalization [106]. More recently, Raza et al. improved the regression-based cell detection framework by introducing a mapping filter and inverting it for reconstruction [107].

Mitosis Detection: Mitosis is a complex process where cells are undergoing various transformations and divide into two daughter cells. As a special case of cell counting, the DL-based mitosis detection can also be achieved with detection-based or regression-based methods. Some groups have also proposed architectures that combine handcrafted features and DL features for mitosis detection to increase the efficiency [105], [108]. The regression-based mitosis detection is similar to the segmentation task, which is a pixel-wise classification [104], [109]. The architectures involved are either the encoder-decoder or CNNs that maintain the same resolution level. Chen et al. convert the problem to a pixel-wise regression problem and utilizes transfer learning in auto-encoders [110]. The detection-based mitosis detection is a two-step process [111]. First, the model localizes the potential mitosis. Then, the cropped image or features from the region are classified into mitosis or not. Compared to regression-based methods, the two-step detection-based methods can differentiate mitosis from hard mimics [112].

IV. LEARNING WITH FEW OR WEAK LABELS

The success of modern DL relies on a huge amount of labeled data. However, unlike fields such as natural image classification in which labels are relatively easy and cheap to get, collecting a locally-annotated WSI dataset can be extremely tedious and expensive. Because of the large volume of whole-slide images, going through the whole image for annotations requires not only expertise in pathology but also much labor. Hence, researchers are interested in improving the performance of deep models by exploiting the unlabeled data or weakly-labeled data. Since most of the works on semi-supervised learning or weakly-supervised learning in WSI are focusing on classification, we assume all the discussion below is default to classification unless specified.

A. Semi-supervised Learning

In semi-supervised learning (SSL), we are given a set of N fully annotated WSIs $X_1, X_2, \dots, X_N \in \mathcal{X}$ with (x_i^j, y_i^j) to be the j^{th} patch of slide X_i labeled as $y_i^j \in \mathcal{Y}$. Additionally, we have M unlabeled slides $X_{N+1}, X_{N+2}, \dots, X_{N+M} \in \mathcal{X}$. The motivation of SSL is how to effectively leverage those N (possibly small) labeled samples and M (usually massive) unlabeled samples to obtain a better predictive model compared to use the N labeled data alone. There are mainly three types of DL-based semi-supervised methods on WSI (**Table III**). The first method is the pseudo-label method,

where unlabelled data is labeled by trained models and added to training set as pseudo-labeled data [118], [119]. Intuitively, the pseudo-label method is a form of Entropy regularization method that favors low-density decision boundary, meaning that data points close to each other in the manifold space are favored to share the same labels. The second method is based on generative adversarial networks (GANs), where a conditional GAN is trained to predict fake/real data as well as class labels [124]. GAN works here because the unlabeled data can help to approximate the true data distribution, hence achieving the similar effect of data augmentation [139]. The third method is based on autoencoders, where data regardless of labeled or not is first fed to an autoencoder to learn a feature embedding, afterwards only the labeled data is used to train the encoder-classifier network [120]–[123]. With the extra unlabeled data, the autoencoder can learn more compact feature representation and leads to better generalization [140].

B. Weakly-supervised Learning

In practice, slide-level labeled WSIs are much easier to obtain, as those data come from routine diagnosis and do not require additional annotations. Learning with only the slide-level label in WSI is commonly known as weakly-supervised learning (WSL), where a whole-slide image is considered to be a "bag" full of "instances" (patches). Formally, we are given a set of N samples $X_1, X_2, \dots, X_N \in \mathcal{X}$ with slide labels $Y_1, Y_2, \dots, Y_N \in \mathcal{Y}$, respectively. Under such "weak" label information, we are trying to predict slide-level outcome and/or patch-level outcome (x_i^j, y_i^j) , where x_i^j is the j^{th} patch of X_i labeled with $y_i^j \in \mathcal{Y}$. There are mainly two types of WSL methods used in WSI: (1) patch aggregation methods, where patch-level information of each WSI is extracted first and then aggregated to encode global context; (2) recurrent attention methods, where the models attend to a series of discriminative regions and makes predictions based on these observations or glimpses. In the following contents, we will discuss both methods in details.

1) Patch Aggregation Methods: The patch aggregation methods follow a bottom-up pipeline and mainly contain two steps. The first step is to obtain patch-level representation for every patch of the input WSI. Secondly, aggregation or voting are employed to obtain a global representation, which is used to make final slide predictions.

There are in general four types of approaches to obtain patch-level representations in WSL, among which two types are categorized as patch-level learning methods and the other two are more direct methods (**Figure 6**). Of the two patch-level learning methods, the most common one is multiple instance learning (MIL) based method, which train a patch-level classifier using only the slide labels. To achieve that, these methods rely on the MIL assumption: a whole-slide image is benign if all of its patches are benign and is malignant if any of its patches are malignant. Therefore, all patches in benign WSIs are labeled as benign. For malignant whole-slide images, we will initialize a patch-level malignant classifier and run through the entire slide, obtaining a malignant probability map. After that, different patch selection strategies are developed to

TABLE III
METHODS FOR LEARNING WITH FEW OR WEAK LABELS IN WSI

Tasks	Methods	Descriptions	Examples
Semi-Supervised Learning			
Patch-level Learning with unlabeled data	Label Mining with Pseudo Label	Iteratively, use trained model to predict unlabeled images and add them to training set	[118], [119]
	Feature Learning with Autoencoders	Train autoencoders on unlabeled set, then use the encoders to extract features and train another tasks-specific classifier	[19], [120]–[123]
	Feature Learning with GANs	Train conditional GANs with an auxiliary task to predict patch labels	[124]
Weakly-Supervised Learning			
Learning with slide labels by Patch Aggregation	MIL-based Patch Learning + Decision-level Fusion	Train patch-level CNNs with variants of MIL assumptions, then aggregate the patch-level predictions by simple pooling, voting or another shallow learners	[11], [125]–[128]
	Unsupervised Feature Learning + Feature-level Fusion	Train feature extractors using unsupervised techniques such as GANs, Autoencoders, Contrastive training and etc., then train another classifier to predict slide labels with those compact patch-level feature representations	[129]–[132]
	Transfer Learning / Customized CNN + Feature-level Fusion	Use a pre-trained network or a self-customized CNN as feature extractor, then aggregation the patch-features and train end-to-end	[133]–[135]
Learning with slide labels by Recurrent Attention	Gaussian Filters Attention + RNN encoders	Recursively select representative tiles through parametrized Gaussian filters and encode the tile information using RNN. When recursion is done, make slide-level predictions based on those representative tiles	[136]
	Deep RL-based Attention + RNN encoders	Use deep reinforcement learning methods to select representative tiles to maximize the prediction accuracy	[137], [138]

propagate slide labels to patches. The most common way is to assign top-K (in terms of prediction confidence) predicted patches as malignant [127]. Other methods utilize the EM-based methods [11], [126] or incorporate the cluster assumption [125], which take advantages of the spatial relation and/or context information in the WSI. Those malignant patches from the malignant slide and the assured benign patches from the benign slide are fed to the classifier for training and the process is iterated until convergence. During training, the classifier has to calibrate itself to predict high probability on malignant patches so that they can be differentiated from the benign patches. Through this training procedure, we can obtain a patch-level classifier, which can produce both the patch-level feature representations or prediction scores for down-stream components.

The second type of patch-level learning methods are feature learning based methods. These types of methods focus on using unsupervised learning techniques to encode patch context efficiently, so as to generate a compact feature vector for each patch. Various types of unsupervised learning methods can be applied here, including:

- Autoencoders [129], [132], where an encoder-decoder network is learnt by reconstructing the input. Then, the encoder alone is used as a feature extractor;
- Contrastive training [129], where the feature extractor is trained to differentiate different input pairs.
- BiGAN [130], where an encoder is trained jointly with a generative adversarial network (GAN).
- Contrastive predictive coding (CPC) [131], where the mutual information between adjacent patches are maximized.

[130] compared the three feature learning methods except CPC and reported that features learnt by BiGAN produced better performance in down-stream task. Note that the feature learning based methods can only generate patch-level feature representations for down-stream components.

For the other two direct methods, the first one is transfer-learning based methods [133], [141], where an image classification network pretrained on other datasets such as Imagenet [6] are used to initialize a patch-level CNN. The second one is simply a randomly-initialized customized CNN [135]. A major difference between direct methods and patch-level learning methods are the former one is trainable end-to-end with the following aggregation components, while the latter is usually fixed and will not update with the following slide-level learning.

Once the patch-level network is obtained, the second major component in the WSL pipeline is information aggregation. There are mainly two types of aggregation methods: (1) decision-level aggregation, where patch-level predictions are aggregated to obtain slide-level predictions; (2) feature-level aggregation, where patch-level features are aggregated to produce a compact global representation.

The general pipeline of decision-level aggregation is shown in **Figure 6**. A patch prediction probability map is first obtained by a full inference pass of the trained patch-level classifier on the slide. Then optionally, statistical features and morphological features can be extracted from this probability map. Finally, the probability map and/or the extracted features are fed into a decision classifier to obtain the slide-level prediction. Different configurations of decision classifiers are widely explored in the community, including:

- Pooling [11], where the slide prediction probability is equal to the maximum (max-pool) or average (average-vote) patch probability in the probability map;
- MIL-based Methods [134], where a slide is classified as negative only if all of the patches are negative;
- Second-level classifier [11], [127], [128], where statistical features such as predicted class histogram and/or morphological features such as sizes of connected components are extracted; then, another classifier such as Logistic

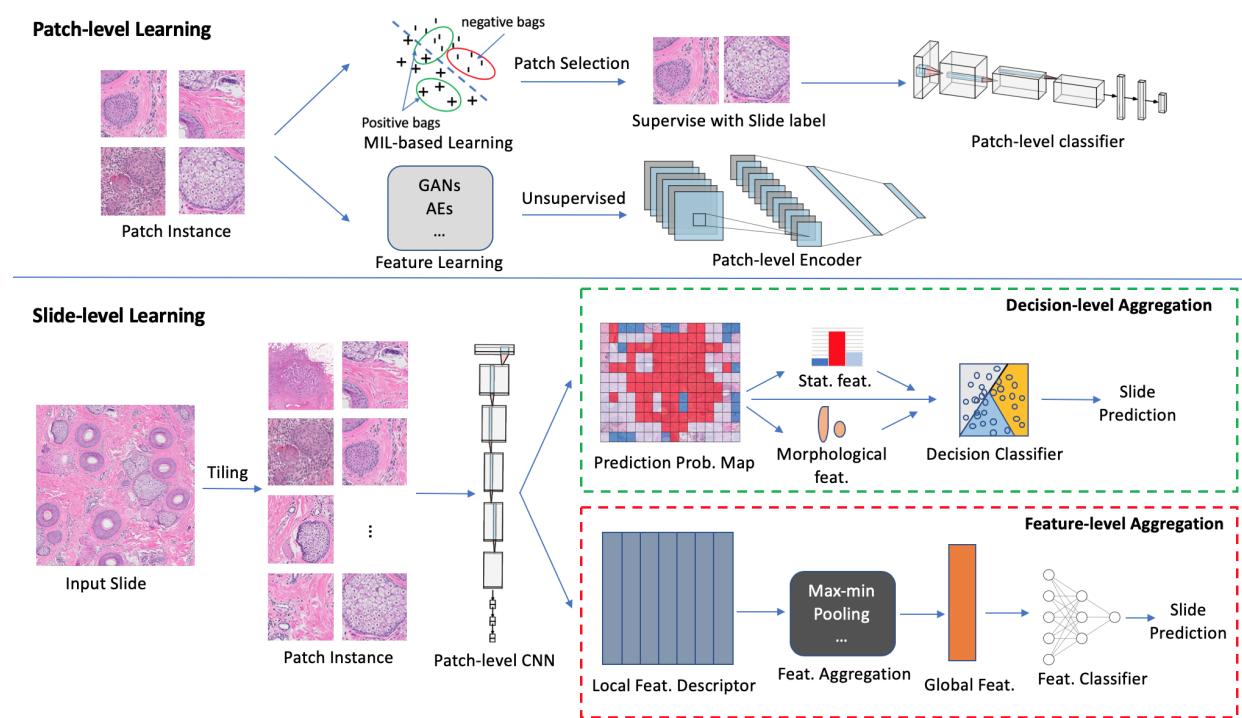


Fig. 6. General pipeline of Patch Aggregation Methods for Weakly-supervised Learning in WSI, which consists of patch-level representation learning and slide-level learning. Top: typical patch-level learning methods in WSL can be classified into MIL-based and feature learning; bottom: depending on different aggregation strategies, slide-level learning methods can be classified into decision-level and feature-level.

Regression, SVM, Tree-based Methods or MLP is trained using the probability histogram and slide labels.

It is often observed that using a second-level classifier in decision-based WSL obtain better performance [11], [127] in trade off with more engineering efforts and computation cost. Note that to use decision-level aggregation, a patch-level classifier has to be trained. Therefore, the only applicable patch-level learning method is the MIL-based method according to the above discussion.

The high-level pipeline of feature-level aggregation methods is shown in the bottom part of **Figure 6**. The feature-level aggregation strategies have been studied extensively. One of the most common and simple strategies is to use mean pooling [133], [142], where local features are averaged to produce global features. This method is fast and usually a very good baseline for most of the applications. Proposed in [143], P-norm Pooling, also known as Softmax Pooling, suggested to amplify the effect of a few most-activated patches to enhance discriminating capability of the global representation. Fisher Vector Encoding [60] is also applied here to obtain a compact representation by using Fisher Vector methods to encode the local features. Max-min Layer [133], where only the top and bottom K most-activated patches are selected to represent the global descriptor was also proven to be effective. [144] proposed block-wise pooling strategies, where local features were first grouped into block and pooling was applied within each block individually. Recently, [127] developed an RNN-based methods for feature aggregation, where patch-level features are fed into a recurrent network sequentially to obtain global representation. Once we have the global representation,

a feature classifier is employed on top of it. Typically, a simple linear layer or an MLP is used here for end-to-end training [127], [133]. Otherwise, a second-level classifier such as SVM or random forest are shown to be more effective in many applications [11], [80], [141].

2) *Recurrent Attention Methods*: The previous patch aggregation methods encode and involve every individual patch to make the final predictions. This is in fact, different to the common diagnosis process of pathologists, where over time fewer fixation and attention is paid to less discriminative regions [145]. Motivating by this observation, [136] made an assumption that the diagnosis of a WSI is based on a few representative tiles and proposed a novel recurrent attention framework for WSL in WSI classification. Basically, the model works in a recurrent fashion. During each step, a discriminative or representative glimpse, which is significantly smaller than the original WSI, is attended and extracted from the WSI. Then, the attended glimpse and the encoded information of historical glimpses are processed by a recurrent network (e.g. LSTM), which generates a new glimpse location. After some fixed recurrent steps, the internal state of the recurrent network is used to predict the slide labels. To make the entire process trainable end-to-end, the glimpse extraction is done by fitting two Gaussian filters, which are then multiplied with the original WSI. Also, by utilizing the MIL assumption, they added an auxiliary glimpse label predictions task to improve the glimpse extraction module, where the glimpse label is simply the slide labels. The intuition here is to encourage the model to attend to regions that share the same slide labels.

Recently, two novel recurrent attention methods are pro-

posed, which both apply deep reinforcement learning (DRL) techniques to guided the glimpse selection [137], [138]. Basically, the localization/attention network is considered as the agent in classical DRL and the WSI is considered as the environment. The agent tries to maximize the reward, which at here is the classification accuracy, by interacting with the environment. Both methods utilize the policy gradient methods to propagate the training signal into the localization network. Additionally, [138] proposed to include the context information to predict glimpse location, by encoding a downsampled version of the original WSI.

The benefit of recurrent attention methods lie in two folds. First, it significantly saves computation power and decrease the inference time, by selectively choosing a few tiles to make predictions. Second, it provides more interpretation and localization by choosing the representative glimpses. We believe this two properties will further advance the development of recurrent attention methods in the application of WSI.

V. MODEL INTERPRETATION FOR WSI

Deep model interpretation aims to provide quantitative measures or visualizations to explain why a model works or fails. It's of great significance for model design, diagnosis, and building trust among users, which is essential for meaningful applications in biomedical areas. The transparency provided by model interpretation can benefit the development of AI in multiple stages from when AI is weaker to when AI is stronger than humans [151].

The medical domain, especially computational digital pathology, requires explanations and insights for a better understanding beyond standard quantitative performance evaluation [155]. Besides model diagnosis, interpretation methods can also help to identify the data set biases in digital pathology, which is one major bottleneck for the success of DL in medical applications [150]. Because of the relatively limited number of data, the complexity, and the heterogeneity by nature, the deep models for WSI have to deal with data set bias, class correlated bias, and sampling bias [150]. First, the data set bias or batch effects can be caused by various platforms or protocols among different institutes. The imbalanced prevalence among different data sets can also contribute to data set bias. These biases typically affect the entire data set and can lead to reduced cross-dataset generalization. Second, the class correlated bias refers to features that are accidentally correlated with labels but without biological meanings. For example, the pen markers on the slide can be an undesired feature if they are correlated with a specific class. Third, the sampling bias is caused by specific procedures when sampling the WSI and preparing the data. The standard quantitative performance evaluation can be hard to capture these biases. However, the model interpretation techniques can serve as a powerful tool to reveal the bias and guide the model design.

Because of the two-stage analysis of WSI, the interpretation is also performed at the patch-level and slide-level respectively.

A. Patch-Level Interpretation

The model interpretation for patch-level images can be directly adopted from techniques developed for natural images.

For image classification task in computer vision, the model interpretation aims to identify class-discriminate information. Various techniques have been developed for visualizing the features and patterns learned by the neural network or locating the regions in an image that contribute most to a specific prediction [146]–[149], [151], [153], [156]. We have summarized these techniques in **Figure 7** and **Table IV**.

The first category of interpretation methods aims to understand the features/patterns learned by the model. One straightforward approach is to generate an image that maximize the activation of a trained network by computing the gradient of the class score with respect to the input image [146]. Besides synthesizing the whole image, researchers also apply deconvolution to reveal the structures/features that excite each network layer [147], [156], which are typically the invariant patterns learned by the neural network after training. For example, Cruz-Roa et al. have trained a model for basal-cell carcinoma detection and interpreted the model by visualizing the features learned by deep neural networks [57].

The second category of interpretation methods focused on locating the regions in a specific image that contributes to the classification given a trained deep model. There are three approaches to realize the pattern localization in the input image for deep neural networks: linear approximation, back propagation, and occlusion sensitivity. Because linear models can easily provide the connections between input features and the classifications, the first localization approach approximates the nonlinear deep neural network with a linear model using methods such as Taylor decomposition and then generate image specific class saliency visualization [146], [148]. However, the linear approximation based visualization methods haven't been successfully applied to deep WSI model interpretation.

The second localization approach takes advantage of the back propagation through the trained network with prediction scores or gradients. The layer-wise relevance propagation (LRP) method generates a pixel-wise explanation heat map by distributing the output value to the input dimension through a backward pass [149]. For example, Hagele et al. have performed patch-level pixel-wise visualization by layer-wise relevance propagation (LRP) and assign a relevance score, which indicates how much the individual pixel contributes to the classifier decision to every pixel in the input image [150]. The gradient-weighted class activation mapping (Grad-CAM) method uses the gradients of any target concept and flow into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept [151]. For example, Korbar et al. have applied the Grad-CAM for visualizing the prediction of the outcomes for colorectal polyps [152].

Lastly, the occlusion sensitivity methods generate region or pixel contribution to the classification by systematically perturbing or erasing a pixel or a region in the input image [147], [153]. The occlusion method is easy to understand but very computationally expensive.

B. Slide-Level Interpretation

The slide level interpretation is unique for WSI in computational digital pathology. Because of the heterogeneity of the

TABLE IV
DL MODEL INTERPRETATION METHODS IN WSI

Tasks	Methods	Descriptions	Examples
Patch-Level Interpretation			
Model-based	Image synthesis [146]	Generates an representative image that maximize the class scores	NA
	Feature visualization [147]	Use deconvolution to reveal the structures that excite the each network layer	[57]
Image-specific	Linear approximation [146], [148]	Obtain a saliency map by Taylor decomposition	NA
	Layer-wise relevance propagation (LRP) [149]	Pixel-wise visualization by distributing the output value to input dimensions through backward pass	[150]
	Gradient-weighted class activation mapping (Grad-CAM) [151]	Coarse localization by gradients back-propagation	[152]
	Occlusion sensitivity [147], [153]	Locating the features by occlusions on the input image	NA
Slide-Level Interpretation			
Slide Heatmaps	Patch-level probability scores	Use patch-level probability scores to generate the slide-level heatmaps	[154]
	Gradient-Based	Backproject the patch-level classification to the input for activation maps	[152]

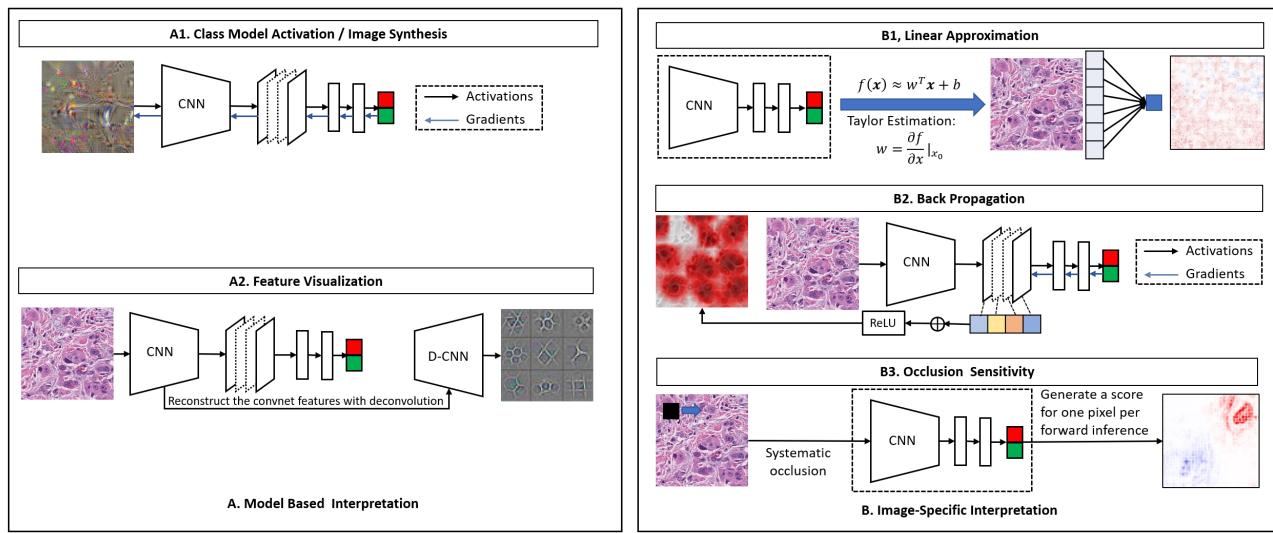


Fig. 7. Patch-level interpretation methods for deep CNN based classification. The interpretation can be classified into two categories: model-based interpretation and image-specific interpretation. The model-based interpretation aims to visualize the patterns/features learned by the model. The image-specific interpretation focuses on locating the regions in the image that contributes to the classification.

tissues in the slide, features extracted from various patches need to be aggregated for slide-level classification. The slide-level decision making is usually interpreted by generating a patch-wise heat maps for the slide. The most straightforward way is to plot the heat map with patch-wise classification probabilities [154]. However, dense sampling the slide to improve the heatmap resolution can be time consuming. One the other hand, we can also project the patch-level classification back to the first layer of the network and obtain the activation map for the entire patch to generate the heatmaps for the slide [152].

VI. MULTI-MODAL DATA INTEGRATION FOR WSI

Multi-modal data integration aims to improve the model performance by combining information extracted from different modalities of the same patient. The main hypothesis is that data from different modalities contain both shared and complementary information. In machine learning, each modality is a view of the subject, and the multi-modal integration is called multi-view learning [157]. In the era of DL, multi-view learning can be approached by joint representation, coordinated representation, and encoder-decoders [158]. The

joint representation projects features from various modalities to a common space and integrate them by concatenation, summation, or multiplication, which utilizes the complementary information from each modality. On the other hand, coordinated representation learns separated but constrained representations for each modality. With proper regularizations, we can utilize the shared information from various modalities by imposing cross-modal similarity constraints. The encoder-decoder framework integrates multi-modal data by mapping one modality into another, which also exploits the shared information and gets rid of discrepancies between modalities.

For multi-modal data integration in WSI, current studies focus on feature concatenation, cross-modality correlation, and cross-modality prediction (**Figure 8**). To combine the complementary information, we first learn feature representations from various modality and then concatenate them for improved prediction performance. To take advantage of the shared information, we can enforce similarity constraints such as correlations or perform cross-modal predictions to get rid of noises.

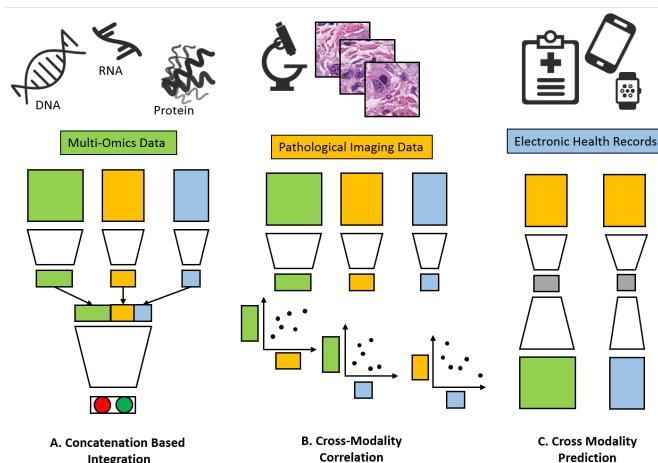


Fig. 8. Multi-modal data integration for whole-slide images. Data from multiple domains (e.g., genomics and electronic health records) can be integrated with WSI pathological features by feature concatenation, cross-modality correlation, and cross-modality prediction.

A. Integration by feature concatenation

Feature concatenation is one of the most straightforward methods for integrating multi-modal biomedical data. We first perform feature extraction for each modality. For some modalities with a massive number of features such as genomics data, we also need to apply feature selection or dimension reduction before concatenating the features. With the availability of DL, we can apply various neural networks to each modality independently for feature representation. After transforming the raw features into a typically low-dimension space, we can integrate the information by concatenation. For example, Mobadersany et al. integrate the intermediate features extracted from whole-slide images with the selected genomics features by concatenation [159]. They feed the concatenated features into a fully connected network for the prediction of cancer survival. Although concatenation-based integration is simple to implement, the interactions among various modalities are not fully employed during the learning process.

B. Integration by cross-modality correlation

With the hypothesis that data from various modalities share common information, we can improve multi-modal feature representation with similarity constraints such as correlation and distance measures. By correlating the extracted WSI features with features extracted from other modalities, we can enhance the feature representation by removing biases. For example, Azuaje et al. correlate the WSI features extracted by CNNs with the selected proteomics features to identify potential biomarkers for kidney cancer [160]. However, the correlation has not been utilized to improve the feature representation for either modality. Thus, further improvements can be DL-based feature representation with correlation constraints.

C. Integration by cross-modality retrieval

Besides the cross-modality correlation, another integration method to take advantage of the shared information is cross-modal retrieval. Cross-modal retrieval aims to retrieve one modality using another modality as the input. The encoder-decoder is one of the most popular structures for cross-modality retrieval. By mapping features from one modality to another, we can extract shared information among modalities and get rid of discrepancies. The cross-modal retrieval studies connect WSI with other modalities by predicting the features of other modalities from WSI. For example, Schmauch et al. predict RNA-seq profiles using whole-slide images [161].

VII. DISCUSSIONS AND FUTURE DIRECTIONS

Beyond the challenges and opportunities briefly summarized in [14], in this article, we provide a comprehensive review of WSI using advanced AI such as DL focusing on three major challenges.

The first major challenge is the lack of labeled WSI for DL training. As a data-driven approach, the performance of DL heavily relies on the amount of well-annotated WSI data. However, the acquisition of labeled WSI data are typically privacy-constrained and the annotation by pathologists is time-consuming. Thus, besides using as much of available dataset as possible, there are a few strategies and opportunities to deal with this challenge. The first strategy is to gather resources for data collection and annotation, and to release the de-identified annotated data to the community for sharing. As shown in **Table V**, there are some public datasets available for sharing as of today. The second strategy is to utilize data augmentation to enlarge the training dataset: besides simple rotation and scaling, advanced data augmentation techniques include GAN-based approaches. The third strategy is to develop advanced semi-supervised and weakly-supervised methods so to utilize the unlabeled data and partially labeled data, which are usually much easier to obtain. The last strategy is to investigate domain adaption approaches so to utilize annotated data from other imaging domains or annotated pathological images of various diseases.

The second major challenge for analyzing WSI using DL is the heterogeneity (i.e. the tissue types and patterns in each slide under various disease conditions are heterogeneous). The highly heterogeneous data make the feature representation for WSIs a very difficult problem. Not only a large amount of data containing the heterogeneous patterns and tissue types are required for the training, advanced DL architectures are also needed. Current studies typically train a separate model for each new image class encountered. If the data is insufficient, researchers explore transfer learning or domain adaption to accelerate the feature representation learning. In analyzing WSI using DL, instead of training a separate model for each disease condition, training a universal feature representation network for the basic tissue types and patterns underlying various disease conditions and tasks. A beginning of the universal feature representation learning is to share data and model, which is shown in **Table V**.

The third major challenge for analyzing WSI using DL lies in model interpretation and transparency. As an application

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
TABLE V
SUMMARY OF SELECTED DATASETS AND TOOLS FOR DEEP LEARNING IN WSI

Dataset		
Dataset Name	Description	Link
The Cancer Genome Atlas (TCGA) [162]	A cancer genomics dataset consisting of 33 cancer types	https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
MITOS-ATYPIA 14 [163]	A breast cancer dataset consisting of 1400 frames of WSIs at different magnification levels	https://mitos-atypia-14.grand-challenge.org/
Gland Segmentation Challenge Contest (GlaS@MICCAI'2015) [164]	A colon adenocarcinoma tissue dataset consisting of 165 WSIs	https://warwick.ac.uk/fac/sci/dcs/research/tia/glascontest/
Tumor Proliferation Assessment Challenge 2016 (TUPAC16) [165]	A breast cancer dataset consisting of 500 WSIs selected from TCGA dataset	http://tupac.tue-image.nl/
CAMELYON16 [166], [167]	A sentinel lymph node dataset consisting of 270 WSIs	https://camelyon16.grand-challenge.org/
CAMELYON17 [167], [168]	A breast cancer (breast lymph node) dataset consisting of 50 WSIs from three centers	https://camelyon17.grand-challenge.org/
BACH ICIAR 2018 [169]	A breast cancer histology dataset consisting of 400 labeled microscopy images and 30 WSIs	https://iciar2018-challenge.grand-challenge.org/
Prostate Fused-MRI-Pathology [170]–[173]	A radical prostatectomy specimens dataset consisting of 114 WSIs	https://wiki.cancerimagingarchive.net/display/Public/Prostate+Fused-MRI-Pathology
Breast Metastases to Auxiliary Lymph Nodes [127]	A breast cancer (breast lymph node) dataset consisting of 130 WSIs from 78 patients	https://wiki.cancerimagingarchive.net/display/Public/Breast+Metastases+to+Auxillary+Lymph+Nodes#4dc5f53338634b35a3500cbef18472e0
Post-NAT-BRCA [174]	A dataset consisting of 96 WSIs taken from patients with residual invasive breast cancer following neoadjuvant therapy	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=52758117
Lung Fused-CT-Pathology [175]	A co-registered lung CT/Pathology dataset consisting of 25 WSIs with extent of invasive Adenocarcinoma	https://wiki.cancerimagingarchive.net/display/Public/Lung+Fused-CT-Pathology
PatchCamelyon (PCam) [176]	A metastatic tissue dataset consisting of 327680 patches (32x32) extracted from histopathologic scans of lymph node sections	https://github.com/basveeling/pcam
Kimia Path24 [177]	A tissue texture dataset consisting of 24 WSIs from different body parts	https://github.com/KimiaLab/KimiaPath24
Breast Cancer Histopathological Database (BreakHis) [67]	A breast cancer dataset consisting of 9,109 WSIs from 82 patients	https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/
Stanford Tissue Microarray database [178]	A tissue microarray dataset consisting of 4328 slides	https://tma.im/cgi-bin/home.pl
Atlas of Digital Pathology (ADP) [179]	A carefully selected tissue dataset consisting of 100 whole-slide images and 57 hierarchical HTTs	http://www.dsp.utoronto.ca/projects/ADP/
Tools		
Tool Name	Tasks	Link
DeepSlide [180]	A sliding window framework for classification of high resolution WSIs, often microscopy or histopathology images	https://github.com/BMIRDS/deepslide
MIL [127]	A Weakly-supervised tile level classifier	https://github.com/MSKCC-Computational-Pathology/MIL-nature-medicine-2019
QuPath [181]	A WSI analysis and digital pathology system that supports visualization, annotation, diseases classification and stain estimation	https://github.com/qupath/qupath

to the medical domain, model interpretation is essential for building trust among users, which is crucial for clinical adoption. To build a transparent deep learning model, researchers have developed various interpretation approaches for WSI, as discussed in **Section V**. However, most of the existing methods focus on only one aspect of the interpretation that cannot provide a comprehensive reasoning process. Thus, novel interpretation methods remain to be developed to clarify the decision-making process of DL models. On the other hand, the fusing of DL features with hand-crafted features can also help the model interpretation for clinical adoption.

In this paper, we have reviewed the DL approaches for WSI analysis including quality control and preprocessing, supervised learning, semi-supervised learning, and weakly supervised learning. We have also reviewed DL model interpretation approaches and multi-modal integration approaches for WSI. Although the adoption of DL for computer-aided

diagnosis using WSI is promising based on the current results, multiple challenges need to be addressed to reach the full potential. By overcoming the challenges such as the lack of labeled data and the lack of model transparency, DL can reshape the diagnosis in computational pathology and lead to precision and personalized computer-aided diagnosis for WSI.

Appendix one text goes here.

ACKNOWLEDGMENT

The authors would like to thank Huiye Liu for her academic suggestions.

REFERENCES

- [1] S. Bainbridge, R. Cake, M. Meredith, P. Furness, and B. Gordon, “Testing times to come? an evaluation of digital pathology capacity across the uk,” *Cancer Research UK. Published November*, 2016.
- [2] A. of American Medical Colleges, *2014 physician specialty data book*. Association of American Medical Colleges, 2015.

- [3] S. Kothari, J. H. Phan, T. H. Stokes, and M. D. Wang, "Pathology imaging informatics for quantitative analysis of whole-slide images," *Journal of the American Medical Informatics Association*, vol. 20, no. 6, pp. 1099–1108, 2013.
- [4] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review," *IEEE reviews in biomedical engineering*, vol. 9, pp. 234–263, 2016.
- [5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [7] T. Mikolov, M. Karafiat, L. Burget, J. Černocký, and S. Khudanpur, "Recurrent neural network based language model," in *Eleventh annual conference of the international speech communication association*, 2010.
- [8] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*, pp. 6645–6649, IEEE, 2013.
- [9] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [10] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, p. 24454, 2016.
- [11] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2424–2433, 2016.
- [12] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records," *Scientific reports*, vol. 6, p. 26094, 2016.
- [13] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of dna-and rna-binding proteins by deep learning," *Nature biotechnology*, vol. 33, no. 8, p. 831, 2015.
- [14] H. R. Tizhoosh and L. Pantanowitz, "Artificial intelligence and digital pathology: Challenges and opportunities," *Journal of pathology informatics*, vol. 9, 2018.
- [15] H. Y. Chang, C. K. Jung, J. I. Woo, S. Lee, J. Cho, S. W. Kim, and T.-Y. Kwak, "Artificial intelligence in pathology," *Journal of pathology and translational medicine*, vol. 53, no. 1, p. 1, 2019.
- [16] D. Tellez, G. Litjens, P. Bandi, W. Bulten, J.-M. Bokhorst, F. Ciompi, and J. van der Laak, "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," *arXiv preprint arXiv:1902.06543*, 2019.
- [17] N. Bayramoglu and J. Heikkilä, "Transfer learning for cell nuclei classification in histopathology images," in *European Conference on Computer Vision*, pp. 532–539, Springer, 2016.
- [18] F. Mahmood, D. Borders, R. Chen, G. N. McKay, K. J. Salimian, A. Baras, and N. J. Durr, "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE transactions on medical imaging*, 2019.
- [19] L. Hou, V. Nguyen, A. B. Kanevsky, D. Samaras, T. M. Kurc, T. Zhao, R. R. Gupta, Y. Gao, W. Chen, D. Foran, et al., "Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images," *Pattern recognition*, vol. 86, pp. 188–200, 2019.
- [20] L. Hou, A. Agarwal, D. Samaras, T. M. Kurc, R. R. Gupta, and J. H. Saltz, "Robust histopathology image analysis: To label or to synthesize?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8533–8542, 2019.
- [21] L. Gupta, B. M. Klinkhammer, P. Boor, D. Merhof, and M. Gadermayr, "Gan-based image enrichment in digital pathology boosts segmentation accuracy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 631–639, Springer, 2019.
- [22] F. Ciompi, O. Geessink, B. E. Bejnordi, G. S. De Souza, A. Baidoshvili, G. Litjens, B. Van Ginneken, I. Nagtegaal, and J. Van Der Laak, "The importance of stain normalization in colorectal tissue classification with convolutional networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 160–163, IEEE, 2017.
- [23] A. Basavanhally and A. Madabhushi, "Em-based segmentation-driven color standardization of digitized histopathology," in *Medical Imaging 2013: Digital Pathology*, vol. 8676, p. 86760G, International Society for Optics and Photonics, 2013.
- [24] B. E. Bejnordi, G. Litjens, N. Timofeeva, I. Otte-Höller, A. Homeyer, N. Karssemeijer, and J. A. van der Laak, "Stain specific standardization of whole-slide histopathological images," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 404–415, 2015.
- [25] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.
- [26] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. van der Laak, et al., "Histopathology stain-color normalization using deep generative models," 2018.
- [27] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013.
- [28] A. Janowczyk, A. Basavanhally, and A. Madabhushi, "Stain normalization using sparse autoencoders (stanosa): Application to digital pathology," *Computerized Medical Imaging and Graphics*, vol. 57, pp. 50–61, 2017.
- [29] A. Ng et al., "Sparse autoencoder," *CS294A Lecture notes*, vol. 72, no. 2011, pp. 1–19, 2011.
- [30] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [31] M. T. Shaban, C. Baur, N. Navab, and S. Albarqouni, "Staingan: Stain style transfer for digital histological images," *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Apr 2019.
- [32] A. Lahiani, J. Gildenblat, I. Klaman, S. Albarqouni, N. Navab, and E. Klaiman, "Virtualization of tissue staining in digital pathology using an unsupervised deep learning approach," in *European Congress on Digital Pathology*, pp. 47–55, Springer, 2019.
- [33] Z. Xu, C. F. Moro, B. Bozóky, and Q. Zhang, "Gan-based virtual restaining: A promising solution for whole slide image analysis," *arXiv preprint arXiv:1901.04059*, 2019.
- [34] E. Yuan and J. Suh, "Neural stain normalization and unsupervised classification of cell nuclei in histopathological breast cancer images," *arXiv preprint arXiv:1811.03815*, 2018.
- [35] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Trans. Graph.*, vol. 35, pp. 110:1–110:11, July 2016.
- [36] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light: Science & Applications*, vol. 8, no. 1, p. 23, 2019.
- [37] N. Bayramoglu, M. Kaakinen, L. Eklund, and J. Heikkila, "Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 64–71, 2017.
- [38] A. Shrivastava, W. Adorno, L. Ehsan, S. A. Ali, S. R. Moore, B. C. Amadi, P. Kelly, D. E. Brown, and S. Syed, "Self-attentive adversarial stain normalization," *arXiv preprint arXiv:1909.01963*, 2019.
- [39] H. Cho, S. Lim, G. Choi, and H. Min, "Neural stain-style transfer learning using gan for histopathological images," *arXiv preprint arXiv:1710.08543*, 2017.
- [40] A. BenTaieb and G. Hamarneh, "Adversarial stain transfer for histopathology image analysis," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 792–802, 2017.
- [41] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Advances in Computer Vision and Pattern Recognition*, p. 189–209, 2017.
- [42] X. Chen, J. Yu, L. Chen, S. Zeng, X. Liu, and S. Cheng, "Multi-stage domain adversarial style reconstruction for cytopathological image stain normalization," *arXiv preprint arXiv:1909.05184*, 2019.
- [43] M. Gadermayr, V. Appel, B. M. Klinkhammer, P. Boor, and D. Merhof, "Which way round? a study on the performance of stain-translation for segmenting arbitrarily dyed histological images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 165–173, Springer, 2018.
- [44] N. Kolkin, J. Salavon, and G. Shakhnarovich, "Style transfer by relaxed optimal transport and self-similarity," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10051–10060, 2019.
- [45] J. Rabin, S. Ferradans, and N. Papadakis, "Adaptive color transfer with relaxed optimal transport," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4852–4856, IEEE, 2014.
- [46] T. de Bel, M. Hermsen, J. Kers, J. van der Laak, and G. Litjens, "Stain-transforming cycle-consistent generative adversarial networks

- for improved segmentation of renal histopathology," in *International Conference on Medical Imaging with Deep Learning*, pp. 151–163, 2019.
- [47] N. Zhou, D. Cai, X. Han, and J. Yao, "Enhanced cycle-consistent generative adversarial network for color normalization of h&e stained images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 694–702, Springer, 2019.
- [48] A. Lahiani, N. Navab, S. Albarqouni, and E. Klaiman, "Perceptual embedding consistency for seamless reconstruction of tilewise style transfer," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Springer International Publishing, 2019.
- [49] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. van der Laak, and P. H. de With, "Stain normalization of histopathology images using generative adversarial networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 573–577, IEEE, 2018.
- [50] M. Babaie and H. R. Tizhoosh, "Deep features for tissue-fold detection in histopathology images," *arXiv preprint arXiv:1903.07011*, 2019.
- [51] L. Mukherjee, A. Keikhosravi, D. Bui, and K. W. Eliceiri, "Convolutional neural networks for whole slide image superresolution," *Biomedical optics express*, vol. 9, no. 11, pp. 5368–5386, 2018.
- [52] U. Upadhyay and S. P. Awate, "Robust super-resolution gan, with manifold-based and perception loss," *arXiv preprint arXiv:1903.06920*, 2019.
- [53] S. Ali, N. K. Alham, C. Verrill, and J. Rittscher, "Ink removal from histopathology whole slide images by combining classification, detection and image generation models," *arXiv preprint arXiv:1905.04385*, 2019.
- [54] S. Kothari, J. H. Phan, T. H. Stokes, A. O. Osunkoya, A. N. Young, and M. D. Wang, "Removing batch effects from histopathological images for enhanced cancer diagnosis," *IEEE journal of biomedical and health informatics*, vol. 18, no. 3, pp. 765–772, 2013.
- [55] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.
- [56] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1868–1873, IEEE, 2017.
- [57] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 403–410, Springer, 2013.
- [58] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, I. Eric, and C. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 1626–1630, IEEE, 2014.
- [59] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Scientific reports*, vol. 7, no. 1, p. 4172, 2017.
- [60] Y. Song, J. J. Zou, H. Chang, and W. Cai, "Adapting fisher vectors for histopathology image classification," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 600–603, IEEE, 2017.
- [61] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, and P. Hufnagl, "Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology," *Computerized Medical Imaging and Graphics*, vol. 61, pp. 2–13, 2017.
- [62] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *Nature medicine*, vol. 24, no. 10, p. 1559, 2018.
- [63] B. Korbar, A. M. Olofson, A. P. Miraflor, C. M. Nicka, M. A. Suriawinata, L. Torresani, A. A. Suriawinata, and S. Hass nanopour, "Deep learning for classification of colorectal polyps on whole-slide images," *Journal of pathology informatics*, vol. 8, 2017.
- [64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [65] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [66] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [67] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *2016 international joint conference on neural networks (IJCNN)*, pp. 2560–2567, IEEE, 2016.
- [68] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in *2016 23rd International conference on pattern recognition (ICPR)*, pp. 2440–2445, IEEE, 2016.
- [69] B. Wei, Z. Han, X. He, and Y. Yin, "Deep learning model based breast cancer histopathological image classification," in *2017 IEEE 2nd international conference on cloud computing and big data analysis (ICCCBDA)*, pp. 348–353, IEEE, 2017.
- [70] N. H. Motlagh, M. Jannesar, H. Abulkheyr, P. Khosravi, O. Elemento, M. Totoni, and I. Hajrasouliha, "Breast cancer histopathological image classification: A deep learning approach," *bioRxiv*, p. 242818, 2018.
- [71] V. Gupta and A. Bhavsar, "Breast cancer histopathological image classification: is magnification important?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 17–24, 2017.
- [72] V. Gupta and A. Bhavsar, "An integrated multi-scale model for breast cancer histopathological image classification using cnn-pooling and color-texture features," in *European Congress on Digital Pathology*, pp. 172–180, Springer, 2019.
- [73] A. Rakhlina, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep convolutional neural networks for breast cancer histology image analysis," in *International Conference Image Analysis and Recognition*, pp. 737–744, Springer, 2018.
- [74] Y. S. Vang, Z. Chen, and X. Xie, "Deep learning framework for multi-class breast cancer histology image classification," in *International Conference Image Analysis and Recognition*, pp. 914–922, Springer, 2018.
- [75] A. Golatkar, D. Anand, and A. Sethi, "Classification of breast cancer histology using deep learning," in *International Conference Image Analysis and Recognition*, pp. 837–844, Springer, 2018.
- [76] V. Gupta and A. Bhavsar, "Sequential modeling of deep features for breast cancer histopathological image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2254–2261, 2018.
- [77] V. Gupta and A. Bhavsar, "Partially-independent framework for breast cancer histopathological image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0, 2019.
- [78] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *Journal of digital imaging*, pp. 1–13, 2019.
- [79] H. Su, F. Liu, Y. Xie, F. Xing, S. Meyyappan, and L. Yang, "Region segmentation in histopathological breast cancer images using deep convolutional neural network," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 55–58, IEEE, 2015.
- [80] Y. Xu, Z. Jia, Y. Ai, F. Zhang, M. Lai, I. Eric, and C. Chang, "Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation," in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 947–951, IEEE, 2015.
- [81] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, 2016.
- [82] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, L. Zhao, and Y. Fan, "Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks," *Neurocomputing*, vol. 229, pp. 88–99, 2017.
- [83] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 933–936, IEEE, 2017.
- [84] Y. Xu, Y. Li, Y. Wang, M. Liu, Y. Fan, M. Lai, I. Eric, and C. Chang, "Gland instance segmentation using deep multichannel neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 12, pp. 2901–2912, 2017.

- [85] X. Xie, Y. Li, M. Zhang, and L. Shen, "Robust segmentation of nucleus in histopathology images via mask r-cnn," in *International MICCAI Brainlesion Workshop*, pp. 428–436, Springer, 2018.
- [86] H. Qu, G. Riedlinger, P. Wu, Q. Huang, J. Yi, S. De, and D. Metaxas, "Joint segmentation and fine-grained classification of nuclei in histopathology images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 900–904, IEEE, 2019.
- [87] S. Öztürk and B. Akdemir, "Cell-type based semantic segmentation of histopathological images using deep convolutional neural networks," *International Journal of Imaging Systems and Technology*, vol. 29, no. 3, pp. 234–246, 2019.
- [88] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [89] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, and J. Hu, "A deep learning algorithm for one-step contour aware nuclei segmentation of histopathological images," *arXiv preprint arXiv:1803.02786*, 2018.
- [90] T. de Bel, M. Hermsen, B. Smeets, L. Hilbrands, J. van der Laak, and G. Litjens, "Automatic segmentation of histopathological slides of renal tissue using deep learning," in *Medical Imaging 2018: Digital Pathology*, vol. 10581, p. 1058112, International Society for Optics and Photonics, 2018.
- [91] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.
- [92] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [93] R. Chen, Y. Jing, and H. Jackson, "Identifying metastases in sentinel lymph nodes with deep convolutional neural networks," *arXiv preprint arXiv:1608.01658*, 2016.
- [94] D. Wang, A. Khosla, R. Gargya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint arXiv:1606.05718*, 2016.
- [95] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, et al., "Detecting cancer metastases on gigapixel pathology images," *arXiv preprint arXiv:1703.02442*, 2017.
- [96] K. Xiao, Z. Wang, T. Xu, and T. Wan, "A deep learning method for detecting and classifying breast cancer metastasis in lymph nodes on histopathological images," *Beijing*, 2017.
- [97] K. Fan, S. Wen, and Z. Deng, "Deep learning for detecting breast cancer metastases on wsi," in *Innovation in Medicine and Healthcare Systems, and Multimedia*, pp. 137–145, Springer, 2019.
- [98] H. H. N. Pham, M. Futakuchi, A. Bychkov, T. Furukawa, K. Kuroda, and J. Fukuoka, "Detection of lung cancer lymph node metastases from whole-slide histopathological images using a two-step deep learning approach," *The American journal of pathology*, 2019.
- [99] H. Lin, H. Chen, Q. Dou, L. Wang, J. Qin, and P.-A. Heng, "Scannet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide image," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 539–546, IEEE, 2018.
- [100] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P.-A. Heng, "Fast scannet: fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1948–1958, 2019.
- [101] K. Paeng, S. Hwang, S. Park, and M. Kim, "A unified framework for tumor proliferation score prediction in breast histopathology," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 231–239, Springer, 2017.
- [102] J. Hung and A. Carpenter, "Applying faster r-cnn for object detection on malaria images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 56–61, 2017.
- [103] P. Kainz, M. Urschler, S. Schulter, P. Wohlhart, and V. Lepetit, "You should use regression to detect cells," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 276–283, Springer, 2015.
- [104] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 411–418, Springer, 2013.
- [105] H. Wang, A. C. Roa, A. N. Basavanhally, H. L. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *Journal of Medical Imaging*, vol. 1, no. 3, p. 034003, 2014.
- [106] Y. Guo, J. Stein, G. Wu, and A. Krishnamurthy, "Sau-net: A universal deep network for cell counting," in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pp. 299–306, ACM, 2019.
- [107] S. E. A. Raza, K. AbdulJabbar, M. Jamal-Hanjani, S. Veeriah, J. Le Quesne, C. Swanton, and Y. Yuan, "Deconvolving convolutional neural network for cell detection," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 891–894, IEEE, 2019.
- [108] M. Saha, C. Chakraborty, and D. Racoceanu, "Efficient deep learning model for mitosis detection using breast histopathology images," *Computerized Medical Imaging and Graphics*, vol. 64, pp. 29–40, 2018.
- [109] S. Öztürk and B. Akdemir, "A convolutional neural network model for semantic segmentation of mitotic events in microscopy images," *Neural Computing and Applications*, vol. 31, no. 8, pp. 3719–3728, 2019.
- [110] H. Chen, X. Wang, and P. A. Heng, "Automated mitosis detection with deep regression networks," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1204–1207, IEEE, 2016.
- [111] C. Li, X. Wang, W. Liu, and L. J. Latecki, "Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks," *Medical image analysis*, vol. 45, pp. 121–133, 2018.
- [112] H. Chen, Q. Dou, X. Wang, J. Qin, and P. A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [113] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
- [114] N. Ing, Z. Ma, J. Li, H. Salemi, C. Arnold, B. S. Knudsen, and A. Gertych, "Semantic segmentation for prostate cancer grading by convolutional neural networks," in *Medical Imaging 2018: Digital Pathology*, vol. 10581, p. 105811B, International Society for Optics and Photonics, 2018.
- [115] Z. Swiderska-Chadaj, T. Markiewicz, J. Gallego, G. Bueno, B. Grala, and M. Lorent, "Deep learning for damaged tissue detection and segmentation in ki-67 brain tumor specimens based on the u-net model," *Bulletin of the Polish Academy of Sciences. Technical Sciences*, vol. 66, no. 6, 2018.
- [116] A. Madabhushi, A. A. C. Roa, and F. Gonzalez, "High-throughput adaptive sampling for whole-slide histopathology image analysis," Mar. 19 2019. US Patent 10,235,755.
- [117] N. Dong, M. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. Xing, "Reinforced auto-zoom net: Towards accurate and fast breast cancer segmentation in whole-slide images," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 317–325, Springer, 2018.
- [118] S. U. Akram, T. Qaiser, S. Graham, J. Kannala, J. Heikkilä, and N. Rajpoot, "Leveraging unlabeled whole-slide-images for mitosis detection," in *Computational Pathology and Ophthalmic Medical Image Analysis*, pp. 69–77, Springer, 2018.
- [119] A. K. Jaiswal, I. Panshin, D. Shulkin, N. Aneja, and S. Abramov, "Semi-supervised learning for cancer detection of lymph node metastases," *arXiv preprint arXiv:1906.09587*, 2019.
- [120] R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, and E. A. M. Janssen, "Multiclass tissue classification of whole-slide histological images using convolutional neural networks," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods, ICPRAM 2019, Prague, Czech Republic, February 19-21, 2019*, pp. 320–327, 2019.
- [121] A. Cruz-Roa, J. Arévalo, A. Basavanhally, A. Madabhushi, and F. González, "A comparative evaluation of supervised and unsupervised representation learning approaches for anaplastic medulloblastoma differentiation," in *10th International Symposium on Medical Information Processing and Analysis*, vol. 9287, p. 92870G, International Society for Optics and Photonics, 2015.
- [122] L. Hou, K. Singh, D. Samaras, T. M. Kurc, Y. Gao, R. J. Seidman, and J. H. Saltz, "Automatic histopathology image analysis with cnns," in *2016 New York Scientific Data Summit (NYSDS)*, pp. 1–6, IEEE, 2016.
- [123] V. Murthy, L. Hou, D. Samaras, T. M. Kurc, and J. H. Saltz, "Center-focusing multi-task cnn with injected features for classification of glioma nuclear images," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 834–841, IEEE, 2017.
- [124] A. Kapil, A. Meier, A. Zuraw, K. E. Steele, M. C. Rebelatto, G. Schmidt, and N. Brieu, "Deep semi supervised generative learning for automated tumor proportion scoring on nsclc tissue needle biopsies," *Scientific reports*, vol. 8, no. 1, p. 17343, 2018.

- [125] Y. Xu, J.-Y. Zhu, I. Eric, C. Chang, M. Lai, and Z. Tu, "Weakly supervised histopathology cancer image segmentation and classification," *Medical image analysis*, vol. 18, no. 3, pp. 591–604, 2014.
- [126] C. Zhang, Y. Song, D. Zhang, S. Liu, M. Chen, and W. Cai, "Whole slide image classification via iterative patch labelling," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1408–1412, IEEE, 2018.
- [127] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. W. K. Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.
- [128] S. Graham, M. Shaban, T. Qaiser, N. A. Koohbanani, S. A. Khurram, and N. Rajpoot, "Classification of lung cancer histology images using patch-level summary statistics," in *Medical Imaging 2018: Digital Pathology*, vol. 10581, p. 1058119, International Society for Optics and Photonics, 2018.
- [129] D. Tellez, J. van der Laak, and F. Ciompi, "Gigapixel whole-slide image classification using unsupervised image compression and contrastive training," 2018.
- [130] D. Tellez, G. Litjens, J. van der Laak, and F. Ciompi, "Neural image compression for gigapixel histopathology image analysis," *IEEE transactions on pattern analysis and machine intelligence*, 2019.
- [131] M. Y. Lu, R. J. Chen, J. Wang, D. Dillon, and F. Mahmood, "Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding," *arXiv preprint arXiv:1910.10825*, 2019.
- [132] Y. Zhu, L. Tong, S. R. Deshpande, and M. D. Wang, "Improved prediction on heart transplant rejection using convolutional autoencoder and multiple instance learning on whole-slide imaging," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–4, IEEE, 2019.
- [133] P. Courtiol, E. W. Tramel, M. Sanselme, and G. Wainrib, "Classification and disease localization in histopathology using only global labels: A weakly-supervised approach," *arXiv preprint arXiv:1802.02212*, 2018.
- [134] G. Campanella, V. W. K. Silva, and T. J. Fuchs, "Terabyte-scale deep multiple instance learning for classification and localization in pathology," *arXiv preprint arXiv:1805.06983*, 2018.
- [135] M. Z. Alom, T. Aspiras, T. M. Taha, V. K. Asari, T. Bowen, D. Billiter, and S. Arkell, "Advanced deep convolutional neural network approaches for digital pathology image analysis: a comprehensive evaluation with different use cases," *arXiv preprint arXiv:1904.09075*, 2019.
- [136] A. BenTaieb and G. Hamarneh, "Predicting cancer with a recurrent visual attention model for histopathology images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 129–137, Springer, 2018.
- [137] A. Momeni, M. Thibault, and O. Gevaert, "Deep recurrent attention models for histopathological image analysis," *BioRxiv*, p. 438341, 2018.
- [138] T. Qaiser and N. M. Rajpoot, "Learning where to see: A novel attention model for automated immunohistochemical scoring," *IEEE transactions on medical imaging*, 2019.
- [139] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- [140] L. Tong, H. Wu, and M. D. Wang, "Caesnet: Convolutional autoencoder based semi-supervised network for improving multiclass classification of endomicroscopic images," *Journal of the American Medical Informatics Association*, vol. 26, no. 11, pp. 1286–1296, 2019.
- [141] R. Awan, N. A. Koohbanani, M. Shaban, A. Lisowska, and N. Rajpoot, "Context-aware learning using transferable features for classification of breast cancer histology images," in *International Conference Image Analysis and Recognition*, pp. 788–795, Springer, 2018.
- [142] J. Rony, S. Belharbi, J. Dolz, I. Ben Ayed, L. McCaffrey, and E. Granger, "Deep weakly-supervised learning methods for classification and localization in histology images: a survey," *arXiv preprint arXiv:1909.03354*, 2019.
- [143] Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, I. Eric, and C. Chang, "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," *BMC bioinformatics*, vol. 18, no. 1, p. 281, 2017.
- [144] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, Q. Huang, M. Cai, and P.-A. Heng, "Weakly supervised learning for whole slide lung cancer image classification," *Medical Imaging with Deep Learning*, 2018.
- [145] T. T. Brunye, P. A. Carney, K. H. Allison, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Eye movements as an index of pathologist visual expertise: a pilot study," *PloS one*, vol. 9, no. 8, p. e103447, 2014.
- [146] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.
- [147] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*, pp. 818–833, Springer, 2014.
- [148] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, "Explaining nonlinear classification decisions with deep taylor decomposition," *Pattern Recognition*, vol. 65, pp. 211–222, 2017.
- [149] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PloS one*, vol. 10, no. 7, p. e0130140, 2015.
- [150] M. Hägele, P. Seegerer, S. Lapuschkin, M. Bockmayr, W. Samek, F. Klauschen, K.-R. Müller, and A. Binder, "Resolving challenges in deep learning-based analyses of histopathological images using explanation methods," *arXiv preprint arXiv:1908.06943*, 2019.
- [151] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.
- [152] B. Korbar, A. M. Olofson, A. P. Miraflor, C. M. Nicka, M. A. Suriawinata, L. Torresani, A. A. Suriawinata, and S. Hassanpour, "Looking under the hood: Deep neural network visualization to interpret whole-slide image analysis outcomes for colorectal polyps," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 69–75, 2017.
- [153] L. M. Zintgraf, T. S. Cohen, T. Adel, and M. Welling, "Visualizing deep neural network decisions: Prediction difference analysis," *arXiv preprint arXiv:1702.04595*, 2017.
- [154] F. Klauschen, K.-R. Müller, A. Binder, M. Bockmayr, M. Hägele, P. Seegerer, S. Wienert, G. Pruner, S. de Maria, S. Badve, et al., "Scoring of tumor-infiltrating lymphocytes: From visual estimation to machine learning," in *Seminars in cancer biology*, vol. 52, pp. 151–157, Elsevier, 2018.
- [155] A. Holzinger, B. Malle, P. Kieseberg, P. M. Roth, H. Müller, R. Reih, and K. Zatloukal, "Towards the augmented pathologist: Challenges of explainable-ai in digital pathology," *arXiv preprint arXiv:1712.06657*, 2017.
- [156] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, "Understanding neural networks through deep visualization," *arXiv preprint arXiv:1506.06579*, 2015.
- [157] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.
- [158] W. Guo, J. Wang, and S. Wang, "Deep multimodal representation learning: A survey," *IEEE Access*, vol. 7, pp. 63373–63394, 2019.
- [159] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. V. Vega, D. J. Brat, and L. A. Cooper, "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 13, pp. E2970–E2979, 2018.
- [160] F. Azuaje, S.-Y. Kim, D. Perez Hernandez, and G. Dittmar, "Connecting histopathology imaging and proteomics in kidney cancer through machine learning," *Journal of Clinical Medicine*, vol. 8, no. 10, p. 1535, 2019.
- [161] B. Schmauch, A. Romagnoni, E. Pronier, C. Saillard, P. Maillé, J. Calderaro, M. Sefta, S. Toldo, T. Clozel, M. Moarii, et al., "Transcriptomic learning for digital pathology," *BioRxiv*, p. 760173, 2019.
- [162] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The cancer genome atlas (tcga): an immeasurable source of knowledge," *Contemporary oncology*, vol. 19, no. 1A, p. A68, 2015.
- [163] L. Roux, "Mitosis atypia 14 grand challenge," 2014.
- [164] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, et al., "Gland segmentation in colon histology images: The glas challenge contest," *Medical image analysis*, vol. 35, pp. 489–502, 2017.
- [165] M. Veta, Y. J. Heng, N. Stathakis, B. E. Bejnordi, F. Beca, T. Wollmann, K. Rohr, M. A. Shah, D. Wang, M. Rousson, et al., "Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge," *Medical image analysis*, vol. 54, pp. 111–121, 2019.
- [166] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson,

- 1 M. Balkenhol, *et al.*, “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.
- 2 [167] G. Litjens, P. Bandi, B. Ehteshami Bejnordi, O. Geessink, M. Balkenhol, P. Bult, A. Halilovic, M. Hermsen, R. van de Loo, R. Vogels, *et al.*, “1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset,” *GigaScience*, vol. 7, no. 6, p. giy065, 2018.
- 3 [168] P. Bandi, O. Geessink, Q. Manson, M. Van Dijk, M. Balkenhol, M. Hermsen, B. E. Bejnordi, B. Lee, K. Paeng, A. Zhong, *et al.*, “From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge,” *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 550–560, 2018.
- 4 [169] G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, *et al.*, “Bach: Grand challenge on breast cancer histology images,” *Medical image analysis*, 2019.
- 5 [170] R. J. Toth, N. Shih, J. E. Tomaszewski, M. D. Feldman, O. Kutter, D. N. Yu, J. C. Paulus Jr, G. Paladini, and A. Madabhushi, “Histostitcher™: An informatics software platform for reconstructing whole-mount prostate histology using the extensible imaging platform framework,” *Journal of pathology informatics*, vol. 5, 2014.
- 6 [171] G. Xiao, B. N. Bloch, J. Chappelow, E. M. Genega, N. M. Rofsky, R. E. Lenkinski, J. Tomaszewski, M. D. Feldman, M. Rosen, and A. Madabhushi, “Determining histology-mri slice correspondences for defining mri-based disease signatures of prostate cancer,” *Computerized Medical Imaging and Graphics*, vol. 35, no. 7-8, pp. 568–578, 2011.
- 7 [172] J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. De-Wolf, and A. Madabhushi, “Elastic registration of multimodal prostate mri and histology via multiattribute combined mutual information,” *Medical Physics*, vol. 38, no. 4, pp. 2005–2018, 2011.
- 8 [173] A. Singanamalli, M. Rusu, R. E. Sparks, N. N. Shih, A. Ziobor, L.-P. Wang, J. Tomaszewski, M. Rosen, M. Feldman, and A. Madabhushi, “Identifying *in vivo* dce mri markers associated with microvessel architecture and gleason grades of prostate cancer,” *Journal of Magnetic Resonance Imaging*, vol. 43, no. 1, pp. 149–158, 2016.
- 9 [174] M. Peikari, S. Salama, S. Nofech-Mozes, and A. L. Martel, “Automatic cellularity assessment from post-treated breast surgical specimens,” *Cytometry Part A*, vol. 91, no. 11, pp. 1078–1087, 2017.
- 10 [175] M. Rusu, P. Rajiah, R. Gilkeson, M. Yang, C. Donatelli, R. Thawani, F. J. Jacono, P. Linden, and A. Madabhushi, “Co-registration of pre-operative ct with *ex vivo* surgically excised ground glass nodules to define spatial extent of invasive adenocarcinoma on *in vivo* imaging: a proof-of-concept study,” *European radiology*, vol. 27, no. 10, pp. 4209–4217, 2017.
- 11 [176] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling, “Rotation equivariant CNNs for digital pathology,” June 2018.
- 12 [177] M. Babaie, S. Kalra, A. Sriram, C. Mitcheltree, S. Zhu, A. Khatami, S. Rahnamayan, and H. R. Tizhoosh, “Classification and Retrieval of Digital Pathology Scans: A New Dataset.”
- 13 [178] R. J. Marinelli, K. Montgomery, C. L. Liu, N. H. Shah, W. Prapong, M. Nitzberg, Z. K. Zachariah, G. J. Sherlock, Y. Natkunam, R. B. West, *et al.*, “The stanford tissue microarray database,” *Nucleic acids research*, vol. 36, no. suppl_1, pp. D871–D877, 2007.
- 14 [179] M. S. Hosseini, L. Chan, G. Tse, M. Tang, J. Deng, S. Norouzi, C. Rowsell, K. N. Plataniotis, and S. Damaskinos, “Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11747–11756, 2019.
- 15 [180] J. W. Wei, L. J. Tafe, Y. A. Linnik, L. J. Vaickus, N. Tomita, and S. Hassanpour, “Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks,” *Scientific reports*, vol. 9, no. 1, p. 3358, 2019.
- 16 [181] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman, *et al.*, “Qupath: Open source software for digital pathology image analysis,” *Scientific reports*, vol. 7, no. 1, p. 16878, 2017.

Michael Shell Biography text here.

PLACE
PHOTO
HERE

John Doe Biography text here.

Jane Doe Biography text here.