Journal of Computer Science

Review

Challenges of Breast Cancer Detection Based on Histopathology Images

[1]Alaa Mohamed Youssef, [2]Aliaa Abdel-Haleim Abdel-Razik Youssif and [1]Wessam El Behaidy

[1]Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt
[2]College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Egypt

Article history
Received: 24-12-2023
Revised: 08-03-2024
Accepted: 15-03-2024

Corresponding Author:
Alaa Mohamed Youssef
Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt
Email: alaa.youssef@cs.mti.edu.eg

Abstract: Artificial Intelligence (AI) is rapidly evolving every day to become increasingly potent and dependable. AI systems are becoming more sophisticated and are being utilized across various domains with the goal of enhancing human existence. Within the healthcare system, artificial intelligence finds application in handling and documenting large volumes of medical data, conducting analyses of healthcare systems, advancing pharmaceutical development, and aiding physicians in decision-making processes. Machines excel over humans in executing repetitive tasks consistently and reliably. In addition, the performance has recently been enhanced by the emersion of deep learning techniques. Breast cancer presents a significant danger to women globally as it reached 25.4% of new cases diagnosed with cancer types. Its danger increases with its ability to spread outside the breast through blood vessels and lymph vessels. The availability of histopathological images and the advancement in AI and machine learning techniques give new horizons for more investigation and studies of breast histopathology images. In this study, we demonstrate the different steps for detecting and classifying breast cancer through a journey from the preparation of breast tissue specimens to classification clarifying the different techniques used. Furthermore, we will discuss the challenges and solutions for histopathology images and the automated systems used.

Keywords: Breast Cancer, Deep Learning Histopathological Images, Multi-Classification, Segmentation, Handcrafted Features, Transfer Learning

Introduction

Cancer ranks among the primary reasons for mortality on a global scale, as it reached nearly 10 million deaths in 2020 (WHO, 2022). Cancer is the transformation of normal cells to malfunctioning behavior that increases rapidly by attacking and destroying another cell. Breast cancer was the top one over other cancer type for women worldwide in 2018 (WCRFI, 2020; Ponraj and Canessane, 2023). The top seven cancer types are shown in Fig. 1.

Numerous types of cancer can be effectively treated with a high likelihood of cure if identified early and managed appropriately. However, cancer diagnosis is a very time-consuming process as the pathologist needs to examine stained specimens mounted on a glass slide through a microscope. The digital scanner for the Whole Slide Image (WSI), opens the doors for computer vision to tackle the analysis of the WSI to reduce diagnosis time. Whole Slide Image (WSI) is a way of laboratory steps to get histopathological images, either by examining them manually under the microscope or by digitalizing the slide with different magnifications by special scanners (Duenweg *et al*., 2023).

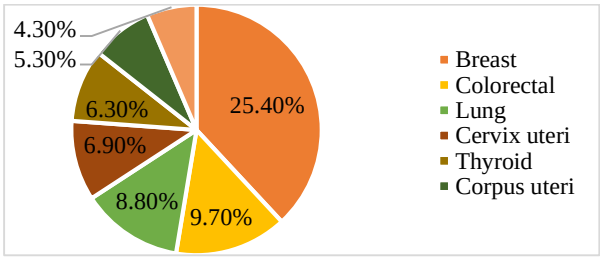

Fig. 1: Cancer type ratio for women worldwide in 2018 (WCRFI, 2020)

Many reviews have discussed the history of histopathological image analysis and the different machine-learning techniques used (Bhargava and Madabhushi, 2016; Shen *et al*., 2017; Gurcan *et al*., 2009; Litjens *et al*., 2017; Xing and Yang, 2016) and we will describe the WSI analysis using different pathology-oriented applications based on machine learning techniques.

The typical steps for the WSI analysis are shown in Fig. 2. The preprocessing steps start with image sampling by taking many patches of small size (e.g., 255×255). These patches may be taken overlapped, randomly, or by mixing between them. Then the feature extraction is performed. Before deep learning, traditional features such as Local Binary Pattern (LBP) and Gray Level Co-occurrence Matrix (GLCM) were extracted. Whereas, deep learning or Convolutional Neural Networks (CNN), optimize the extracted features and classifiers simultaneously, surpassing traditional features in histopathological image analysis (Pankaja and Sivagami, 2024; Hou *et al*., 2016; Xu *et al*., 2016; Sheikhzadeh *et al*., 2018; Litjens *et al*., 2016).

The last step in WSI analysis is the classification of patches as cancer or normal. The classification is performed via suitable machine-learning techniques. Machine Learning (ML) is categorized into two core types; supervised and unsupervised learning. Additionally, there are derivatives conducted from these two core types, such as multiple-instance learning and semi-supervised learning. In this review, we demonstrate the different steps for detecting and classifying breast cancer through a journey from the preparation of breast tissue specimens to classification clarifying the different techniques used. Furthermore, we will explore the challenges and solutions of histopathology images and the automated systems used.

Tissue Preparation and Datasets

This section clarifies the typical laboratory workflow to prepare the tissue, staining processes, and histological slide digitization. In addition to histopathology datasets that may be used for breast cancer classification.

Image Acquisition and Imaging

The typical steps for Whole Slide Images (WSI) image start by taking a small tissue from a suspicious area of the breast by Fine Needle Aspiration (FNA). Then, the biopsy specimen (i.e., a small tissue) is molded in a wax cube at the laboratory, as in step 1 in Fig. 3. The wax cube is sliced into patches using a microtome, as in step 2 in Fig. 3. The average slice thickness is 3-5 µm (Veta *et al*., 2014). Each wax slice is mounted on glass slides from which the WSI is performed, as in step 3 in Fig. 3. However, the slices need to be stained to get clear visualizations in Fig. . 3.

Hematoxylin and Eosin (H and E) stain stands as the most widely employed stain in medical diagnoses. H and E stains produce blue, violet-red stains to biopsy tissue. WSI images always vary due to laboratory stain protocol, staining process, and brightness of the scanner. Finally, a special gigapixel scanners, WSI scanners, are used. This scanner automates all the scanning processes including feeding the scanning platform with the slides, detecting, focusing on relevant tissue areas, acquiring images, compressing, storing, and recordkeeping on a laboratory information system (Veta *et al*., 2014). WSI scanner can achieve spatial resolution with 0.25 µm/pixel by 40X magnification. The WSI images are saved in a multiscale pyramid resolution structure in Fig. . 4.

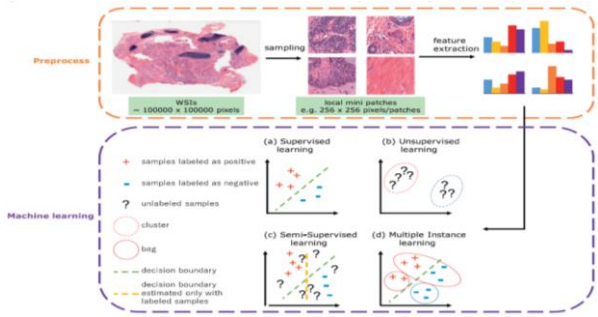

Fig. 2: Digital pathological image analysis steps (Sheikhzadeh *et al*., 2018)

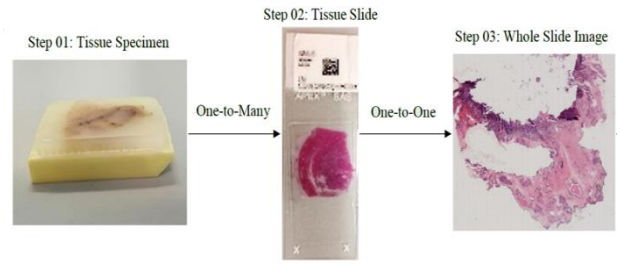

Fig. 3: The process of acquiring WSI images (Dimitriou *et al*., 2019; Wang *et al*., 2012)

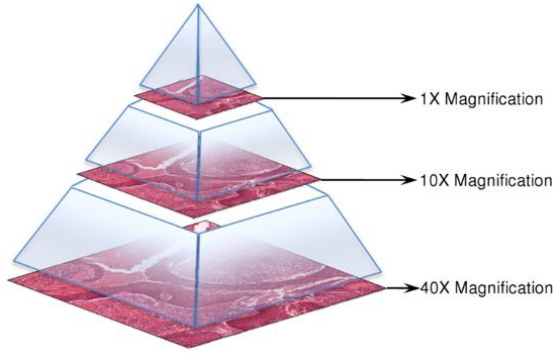

Fig. 4: Typical WSI image with multiscale pyramid resolution (Wang *et al*., 2012)

Datasets

The emersion of machine learning techniques in WSI analysis creates the need to collect histopathology datasets. This need was developed by many medical imaging conferences and workshops since 2007 (Yan *et al*., 2020). The public datasets are used as benchmarks, which introduce a fair metric for comparing the published works as it gives precise task definition and assessment metrics.

For breast cancer, there are small datasets that provide the four main classes: (Normal tissue, benign, in situ, and invasive carcinoma) for automatic classification such as Bioimaging (Bioimaging, 2015) and BACH (Aresta *et al*., 2019). Bioimaging (2015) contains 249 WSI images, which it extended to 3771 breast cancer images in cooperation with Peking University International Hospital (Yan *et al*., 2020). The grand challenge on breast cancer histology images (BACH) (Aresta *et al*., 2019) provides 400 WSI images divided by 100 images per class. In addition, there is a breast cancer dataset that provides two classes (normal-cancerous) for classification such as CAMELYON 17 (Litjens *et al*., 2018) dataset. It contains 681 color images with large size (200,000×100,000) pixels which will need special tools to work on it.

Nevertheless, the appearance of deep learning creates the need for large datasets to build efficient models. For that, large-scale general high-resolution histopathological WSI images are collected such as (TCGA) (Saha *et al*., 2017) and (GTEx) (Sathish *et al*., 2017; Selvi and Suganthi, 2018). The Cancer Genome Atlas (TCGA) (Saha *et al*., 2017) includes more than 10,000 WSIs from several cancer types, whereas Genotype-Tissue Expression (GTEx) (Sathish *et al*., 2017; Selvi and Suganthi, 2018) includes more than 20,000 WSIs from several tissues. The advantage of these datasets is that they provide a genomic profile, which can be used to build a relationship between genome type and morphology.

Materials and Methods

In the following section, the typically basic stages for analyzing the WSI are discussed.

Preprocessing Stage

This stage is responsible for enhancing the quality of the images by manipulating noises and adequate contrast which improves feature extraction. This is performed by using the spatial domain and frequency domain. The spatial domain works on image pixels directly, while the frequency domain works on the Fourier transform of the image (Krithiga and Geetha, 2021). This improvement may look's not significant to the human eye but it enhances the automation process significantly. Furthermore, image preprocessing could be a simple operation such as resizing or converting images to grayscale to reduce the computation time.

Segmentation Stage

Segmentation is the core part of histopathology image analysis in different applications. It extracts the Region of Interest (ROI) and neglects regions with less information. Some segmentation techniques use bottom-up manner like region growth (Pan *et al*., 2006) and Watershed segmentation (Alsubaie *et al*., 2018) to extract morphometric features and determine nuclear pleomorphism grade, identify lymphocyte infiltration, malignancy detection and assess tubule formation such in (Dimopoulos *et al*., 2014; Hu *et al*., 2018; Tosta *et al*., 2017; Vink *et al*., 2013). There is an application that works on high-dimensional histopathological images that uses two stages of segmentation to extract cellular carcinoma structures such as in (Albayrak and Bilgin, 2019; Bejnordi *et al*., 2015a; Jia *et al*., 2017).

Feature Extraction Stage

Histological images require a representation characteristic of tissues and tumor cells in a measurable manner De Matos *et al*., 2019; Das *et al*., 2020). The measurable evaluation of tissue and organ function depends on capturing appropriate features to describe cellular and tissue structures accurately. Feature extraction is the operation of image reduction to get a compressed feature vector. The extracted features need to be sufficiently distinct and identifiable to automatically classify tissues as either normal or malignant and assign corresponding grades (Das *et al*., 2020).

The feature extraction may be a handcrafted feature where the most important feature is extracted like texture, color, and shape of breast cancer histopathological images. However, cancerous cells and normal cells need other special features that can be automatically extracted using deep learning techniques (Krithiga and Geetha, 2021).

Shape-Based Features

The cell shape is a significant feature in detecting cancer cells and the cancer degree. The most significant shape (or morphometric) features for the cell are nuclei area, convex area, and outline. Whereas, the morphometric features of cancer cells are dark nucleoli, prominent, cytoplasm scarcity, cells abnormal growth, an erratic organization of chromosomes, and non-uniform cell size and shape (Lee *et al*., 2023). The grade progression is increased by using the mean values of these features for the nuclear and cell outline as shown in Dundar *et al*. (2011); Kolarević *et al*. (2018); Prvulović *et al*. (2010).

The shape-based analysis can distinguish between 55 breast diseases, 62 carcinoma cases, and 7 Atypical Ductal Hyperplasia (ADH) with WSI taken by Fine Needle Aspiration Cytology (FNAC), as in Niwas *et al*. (2010).

The Hu-moment invariant (Žunić and Žunić, 2014) is a widely utilized global shape feature descriptor for

extracting cell features, irrespective of the cell's orientation, size, and position, as well as its ability to learn new patterns. Hu-moment produces a 7-d feature vector from the image.

Knowing the cancer degree is crucial for devising treatment strategies and predicting patient prognosis in specific types of cancer, including primary brain tumors, soft tissue sarcoma, as well as prostate and breast cancer.

Color-Intensity Based Features

This feature helps in determining the color distribution of the stain of WSI. The main idea is to keep tracking the repetition of each color, as in Li and Plataniotis (2015). However, this feature alone is not enough for detection as there are many variable factors that can affect the color distribution. Nevertheless, this feature can give great add value for image preprocessing and stain color normalization (Sukumar and Gnanamurthy, 2016).

Textural Based Features

Texture refers to the characteristic quality of an image that conveys details about the surface and visual attributes of objects depicted within the image. The texture has two categories regular and stochastic based on the degree of randomness. In a regular texture, the elementary components of an object are arranged periodically, whereas in a stochastic texture, these components are organized randomly.

The histopathological images fall into the category of stochastic texture because of the random distribution of cells. The Gray-Level Cooccurrence Matrix (GLCM) is employed for extracting second-order statistical texture features, calculating these features based on the spatial relationship of pixels (Kazmar *et al*., 2010).

Automatic Features Extraction

CNN convolutional base layers can be used as an automatic feature extractor. CNN Convolutional base consists of several convolutional layers, sub-sampling layers, and activation layers subsequently, that subsequently extract important features from the image. The convolutional layer is responsible for parameter sharing and local connectivity by defining important parameters like the stride value, filter size, use of zero-padding, and number of channels. The sub-sampling layers are responsible for decreasing the computations number and learning parameters and reducing overfitting by reducing the spatial size of the network. while the activation layers activate a few nodes per time to increase computation efficiency (Zhang *et al*., 1990). Deep learning has proven in state of art studies that it has more accurate results (Vink *et al*., 2013).

Feature Selection Stage

Feature selection involves identifying the most impactful feature from a high-dimensional feature space. Removing irrelevant information from high-dimensional input results in reduced overall training time while preserving the original accuracy (Saha *et al*., 2017). The feature selection method primarily relies on evaluating the performance of every potential combination of different features and subsequently choosing the optimal combination based on the results (Salman *et al*., 2014).

Choosing those distinctive features is more essential than choosing the classifier itself. The classification accuracy will be enhanced by increasing the number of selected features from all features (Ponraj and Canessane, 2023; Wang *et al*., 2014). The most popular feature selection strategies:

1) Filter; execute various combinations of statistical features to identify the subset of features that yield the most accurate predictive analysis. The Chi-squared (Jin *et al*., 2006) can be considered as an example, which tests information gain and correlation coefficient scores. As in (Wang *et al*., 2012) they employed a predictive model to assess a combination of features and assign a score based on the accuracy of the model
2) Wrapper; manipulate features selection set as a search problem involves evaluating and comparing different combinations to other combinations. (Wang *et al*., 2016a) they employed a Chain-like Agent Genetic Algorithm (CAGA) to acquire the optimal subset of features for the SVM classifier
3) Embedded; identify which features contribute most effectively to the accuracy of the model during its creation process. Yuan *et al*. (2019), used the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression model for detecting breast cancer

Classification Stage

Classifiers are the algorithms that can teach computers how to distinguish objects based on a set of features. Based on the available data, traditional classifiers can be supervised for the labeled data, unsupervised for the unlabeled data, and semi-supervised for mixed data (Vink *et al*., 2013; Shin *et al*., 2016).

The supervised classifier's goal is to train a model from labeled data of WSI that can infer input image to its labeled class, (Vang *et al*., 2018; Sharma and Mehra, 2020; Zalloum *et al*., 2022). The commonly utilized classifiers are random forest and support vector machine. Whereas, the unsupervised classifier's goal is to train a model that can distinct the hidden structure of unlabeled data and put them in clusters, (Lee *et al*., 2020). The used algorithms are K-mean, autoencoders, and Principal Component Analysis (PCA).

WSI Automated Systems

In this section, automated systems for WSI will be demonstrated.

Computer-Assisted Diagnoses

Computer-Assisted Diagnoses (CAD) (Cicerone and Camp Jr, 2019) are the most recent research interest, as their main goal is to do the diagnosis task of the pathologist. It depends on supervised learning to process WSI and infer the disease category. CAD has proved that it has the same or higher accuracy outcome than manual diagnosis and in less time as the pathologist doesn't search each inch in the WSI manually for the infected parts (Litjens *et al*., 2016).

Additional tasks in diagnosis encompass identifying or delineating the Region of Interest (ROI), such as tumor regions within Whole Slide Images (WSI) (Spanhol *et al*., 2016; Kieffer *et al*., 2017), assessing immunostaining scores (Sheikhzadeh *et al*., 2018; Mungle *et al*., 2017), determining cancer staging (Wang *et al*., 2016b-2015), detecting mitosis (Shah *et al*., 2017; Ludovic *et al*., 2013), segmenting glands Chen *et al*., 2016; Gertych *et al*., 2015; Sirinukunwattana *et al*., 2017) and detection and quantification of vascular invasion (Caie *et al*., 2014).

Content-Based Image Retrieval

The Content Based Image Retrieval (CBIR) (Latif *et al*., 2019; Sridhar *et al*., 2015) is an unsupervised machine learning model that inquires input images to retrieve similar images. In digital pathology, this system is used in diagnosis where professional pathologists can inquire about rare cases. In addition, it is used in education and research where novice pathologists can retrieve similar cases or images of the tissue.

In the CBIR system, a high-speed search is required in addition to its accuracy. Many different techniques have been used to reduce image feature dimensions like hashing (Sparks and Madabhushi, 2016) used for high-speed search and principal component analysis and fast approximate nearest neighbor search such as K-D tree (Medjahed *et al*., 2013).

Histopathological Image Analysis Problems and Solutions

This section will explore the characteristics of pathological images and the challenges that may face machine learning.

Gigantic Image Size

Natural images, like cats or dog images, can keep their features after resizing to 256×256 as an input for the deep learning model. However, the pathological image will lose its feature distinction with resizing and the cell texture has no orientation. This problem was overcome by dividing the single WSI image into patches, each patch is processed individually then the ROI can be extracted from the opposition of cancer patches (Dimitriou *et al*., 2019). There are also wide horizons in this area on how to integrate these patches, make accurate decisions, and avoid the outlier patches.

Furthermore, the computation power of CPU, GPU, and memory has significantly improved in the last decade allowing to processing of large patches (960×960) with more accuracy (Krithiga and Geetha, 2021).

Insufficient Labeled Images

This is a challenge for pathological images as it is existing in small numbers. In addition, it needs to be labeled accurately at patch level and pixel level by professional pathologists in relatively very large WSI slides. On the other hand, general image labeling can be collected and labeled by normal people and there are existing large datasets such as ImageNet (Russakovsky *et al*., 2015).

There are researchers have reused trained models on public image datasets like ImageNet, as in Gutman *et al*. (2016), and have shown significant results. Other researchers have used the available specific dataset but its main drawback is that it is used for specific diseases with specific conditions like magnification level, stain type, image resolution, etc.

The researchers handle insufficient labeled images by one of the following approaches.

Efficient Expansion of Label Data

Pathologists have used different approaches that speed up the labeling process and annotation. One of these approaches is surrounding the ROI parts and applying an image retrieval algorithm to them (Huang and Racoceanu, 2018). Other approaches are also used as eye movement tracing (Brunye *et al*., 2014), mouse cursor tracing (Raghunath *et al*., 2012), and analysis for pathologist viewport (histogram of time elapsed, screen coordinates and zooming levels) (Mercan *et al*., 2016). However, extracting ROI by tracking is not always accurate to the exact boundaries.

On the other hand, active learning is used as an effective approach by many researchers Nalisnik *et al*. (2017); Doyle *et al*. (2011); Padmanabhan *et al*. (2014); Zhu *et al*. (2014); Xu *et al*. (2014); Wang *et al*. (2016a). Active learning is a supervised approach that utilizes a few labeled images to increase discrimination performance by choosing the most valuable unlabeled samples to be labeled. The implementation of this approach was applied by variance reduction (Padmanabhan *et al*., 2014), query-by-committee (Doyle *et al*., 2011), and hypothesis space reduction (Zhu *et al*., 2014) to extract valuable unlabeled samples.

Utilization of Unlabeled Information or Weak Label

Even if we do have not the exact ROI position, the information assigned with WSI of the existence of cancer

or not can be helpful if it is used as a weak label. The main idea of a weak label is that it divides each WSI into patches to be like a "bag" of numerous patches in machine learning settings. A Whole Slide Image (WSI) is classified as cancerous if it contains at least one cancer tissue patch; otherwise, it is labeled as normal. This introduces the problem of weakly-supervised learning (Xu *et al*., 2014; Jia *et al*., 2017) or multiple instance problems (Xu *et al*., 2017a; Dietterich *et al*., 1997).

The weakly supervised learning (Sparks and Madabhushi, 2016; Peikari *et al*., 2015; Miyato *et al*., 2018; Rasmus *et al*., 2015) utilized both labeled and unlabeled data. The real distribution of labeled data is estimated using the unlabeled data. Particular effectiveness is observed when samples in the same class form a well-discriminative cluster.

The multi-instance learning utilizes labeled bags. A method applied to histopathological image analysis includes a support vector machine-based approach (BenTaieb *et al*., 2017) a boosting-based approach (Salman *et al*., 2014), and a deep learning-based approach (Jia *et al*., 2017).

Utilization by Reusing Parameters by Transfer Learning

Machine transfer learning can overcome the limitation posed by the small size of the available dataset. In this methodology, instead of training the model on the dataset from scratch a modified pretrained model for a similar task was used and optimized the parameters for histopathology images. For CNN networks, this is typically done by replacing the network's final three fully connected layers. The parameters of the old trained model can be used without any modification (Xu *et al*., 2017b). It also can be used as initial parameters then we retrain the part or full layers with the new data as in Kieffer *et al*. (2017); Xu *et al*. (2017a); Bayramoglu and Heikkilä (2016); Han *et al*. (2017); Song *et al*. (2015); Romo *et al*. (2014); Doyle *et al*. (2006); Kather *et al*. (2016); Rexhepaj *et al*. (2013); Doyle *et al*. (2007); Linder *et al*. (2012); Wang *et al*. (2017); Bejnordi *et al*. (2015b); Yan *et al*. (2020); Vang *et al*. (2018); Kohl *et al*. (2018); Vesal *et al*. (2018); Sharma and Mehra (2020); Celik *et al*. (2020); Yari *et al*. (2020); Zainudin *et al*. (2020); Burçak *et al*. (2021); Hameed *et al*. (2020); Wadhwa and Kaur (2020).

Many recent researches have proven that transfer learning for CNN model results exceeded the trained model from scratch (Kohl *et al*., 2018).

Different Levels of Magnification Yield Differing Levels of Information

Pathologist WSI diagnosis typically tests the tissue structure under different magnifications. Each magnification level shows a certain structure that infers different information in Fig. . 5.

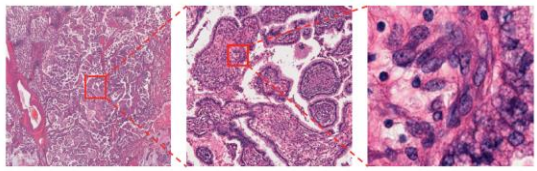

Fig. 5: Histopathological image with different magnification levels, the left side shows low magnification power and the right image shows high magnification power. The image was taken from TCGA (Albayrak and Bilgin, 2019)

There are researchers have utilized images at different magnification levels (Song *et al*., 2015; Romo *et al*., 2014; Doyle *et al*., 2006). As discussed before, processing WSI with its original resolution is hard. The resizing or dividing into patches is relevant to the magnification level and there is an argument on which magnification level is the most informative for diagnosis (Wang *et al*., 2016a; Liu *et al*., 2017).

WSI as Order-Less Texture-Like Image

From the CNN training model point of view, nature images differ from histopathological images. Natural images are represented as objects with orientation and color, whereas histopathological images are represented as patterns or textures. There have been methods that intensively utilize texture structure, such as the Gabor filter bank (Sharma and Mehra, 2020), local binary pattern (Saito *et al*., 2016), and gray level co-occurrence matrix (Ojala *et al*., 1996) and whereas, CNN can be trained on texture using data augmentation by shift and rotate, for example. In addition, a deep texture representation using a CNN is recently developed (Wang *et al*., 2017; Lin *et al*., 2015). A layer of CNN correlation matrix of features is used to get deep texture representation regarding the variance of cell position, with no constraint on WSI image size.

Color Variation and Artifacts

WSI preparation is not an easy straightforward process. The steps differ from one laboratory to another as they may follow different protocols. These protocols include the used stain material. Since most of the steps are done manually, there are undesired effects that may affect the WSI image as the slice may bent, or wrinkled, different thicknesses may lead to blurred parts, and may the slice be marked by a color marker (i.e., Marker pen), Fig. 6. Since these artifacts may affect the result, there are specific algorithm to detect blurring area (Wu *et al*., 2015) and tissue fold (Kothari *et al*., 2013) used as preprocessing stage.

Color variation (Srinidhi *et al*., 2021) is another significant artifact in Fig. . 7. This difference depends on the stain manufacture, staining time condition tissue thickness, and scanning models. If there is a dataset that has sufficient images from all color variations, this will not become a problem for CNN training, but such a dataset is not available till now.

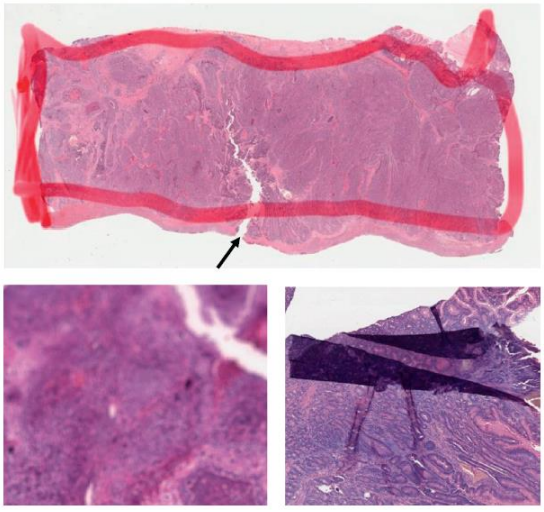

Fig. 6: Different artifacts, the top image is marked by a color marker, the bottom left has blur artifacts, and the bottom right are folded image

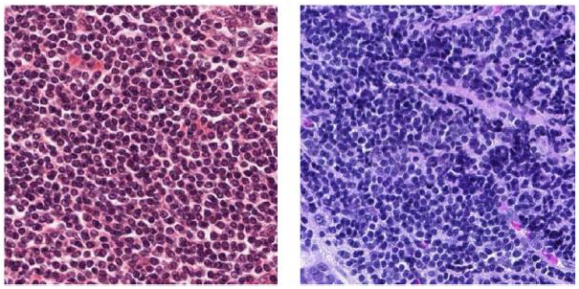

Fig. 7: Color variation of WSI due to two stain

Many researchers have done preprocessing algorithms to avoid this issue. Converting images to grayscale, performing color normalization, and applying color augmentation (Lafarge *et al*., 2017; Lin *et al*., 2018). Gray scale sounds easy implementation but it is not recommended because it will lose a lot of important information. While color normalization tries to normalize the color-magnitude and distribution to be like the trained CNN model, which is more reliable for WSI diagnosis tasks.

Color augmentation (Srinidhi *et al*., 2021) is a process of applying random values on brightness, contrast, hue, and saturation. The main advantage is that it has easy implementation and is suitable for WSI with small color variation. Excessive utilization of color augmentation may result in a distortion of color information, causing the classifier to lose the distinction between objects, so it may be better if combined with color normalization.

Results and Discussion

This survey mainly focused on breast cancer diagnosis using histopathological images. Recently, CNN has proven its effectiveness in feature extraction and classification in breast classification (Krithiga and Geetha, 2021; Kohl *et al*., 2018). For that, we gathered the recent researches that had worked on breast cancer classification using CNN, either binary or multi-classification. The binary classification researches are shown in Table 1, whereas the multi-classification researches are shown in Table 2.

The used CNN models had been modified and tuned to adapt the new features of WSI, using transfer learning. Transfer learning achieved more accurate results than training the CNN from scratch (Kohl *et al*., 2018). Therefore, different authors tried a combination of different classifiers in parallel and then used a voting technique in order to get the best results.

Table 1: The breast cancer research with binary class classification using CNN techniques

Ref. Number	Task	Method	Classifier	Dataset
Kohl *et al*. (2018)	Tumor segmentation	VGG-161 based on ImageNet	Fully connected CNN layer	BACH 2018 and CAMELYON
Sharma and Mehra 2020	Segmentation WSI to normal and infected	VGG-16	SVM	BreakHis
Celik *et al*. (2020)	Segmentation WSI to Normal and infected	DenseNet ResNet-50	Fully connected CNN layer	IDC
Yari *et al*. (2020)	Binary and multiclass classification	DenseNet-121 ResNet-50	Fully connected layer (FC-8)	BreakHis
Zainudin *et al*. (2020)	Binary classification	6 ,13 ,17 and 19 Layers CNN	Soft max layer	MITOS ATYPHIA
Burçak *et al*. 2021)	Binary classification	Custom model	Soft max layer	BreakHis
Hameed *et al*. (2020)	Binary classification	Ensemble VGG-16 and VGG-19	Soft max layer	Custom dataset with 544 WSI Hameed *et al*. (2020)
Wadhwa and Kaur (2020)	Binary classification	DenseNet-201	Fine-tuned CNN Layers	BreakHis

Table 2: The breast cancer research with multi-class classification using CNN techniques

Ref. Number	Task	Method	Classifier	Dataset
Vang *et al*. (2018)	Multi-class classification for WSI	Inception V3	• Majority Voting • Gradient Boosting Machine (GBM) • Logistic Regression (LR)	ICIAR 2018
Vesal *et al*. (2018)	Multi-class classification for WSI	ResNet50 based on ImageNet	Softmax Classifier	BACH 2018
Yan *et al*. (2020)	Multi-class classification for WSI	Inception V3 as Feature extraction	LSTM for fuse patches features and gives final classification	Extended Bioimaging (2015) Yan *et al*. (2020)
Yari *et al*. (2020)	Binary and multiclass classification	DenseNet-121 ResNet-50	Fully Connected layer (FC-8)	BreakHis

Conclusion

In this survey, we provided an intensive study on breast cancer diagnosis based on histopathology image analysis. We had mentioned the way of WSI preparation and different available datasets. Furthermore, we provided an overview of different methodologies used for preprocessing, segmentation, feature extraction, feature selection, and classification. Then, the main WSI analysis challenges and difficulties are discussed in detail and the methods to overcome them. Finally, an intensive study of WSI analysis for breast cancer using various deep-learning CNN networks is discussed.

Acknowledgment

The authors extend their heartfelt appreciation to the editors for their diligent handling of the manuscript, as well as to all reviewers for their valuable and constructive feedback, which has enriched the original submission.

Funding Information

The authors have not received any financial support or funding to report.

Author's Contributions

Alaa Mohamed Youssef: Participated in collected papers, data analysis and written the manuscript.

Aliaa Abdel-Haleim Abdel-Razik Youssif: Participated in reviewed the manuscript.

Wessam El Behaidy: Participated in planned, organizing reviewed the manuscript.

Ethics

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Albayrak, A., & Bilgin, G. (2019). Automatic cell segmentation in histopathological images via two-staged superpixel-based algorithms. *Medical and Biological Engineering and Computing*, *57*, 653-665. https://doi.org/10.1007/978-3-319-06593-9_26

Alsubaie, N., Sirinukunwattana, K., Raza, S. E. A., Snead, D., & Rajpoot, N. (2018, March). A bottom-up approach for tumour differentiation in whole slide images of lung adenocarcinoma. In *Medical Imaging 2018: Digital Pathology 10581*, 104-113. SPIE. https://doi.org/10.1117/12.2293316

Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S. S., Safwan, M., Alex, V., ... & Aguiar, P. (2019). BACH: Grand challenge on breast cancer histology images. *Medical Image Analysis*, *56*, 122-139. https://doi.org/10.1016/j.media.2019.05.010

Bayramoglu, N., & Heikkilä, J. (2016). Transfer learning for cell nuclei classification in histopathology images. *In Computer Vision-ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, (pp. 532-539). Springer International Publishing. https://doi.org/10.1007/978-3-319-49409-8_46

Bejnordi, B. E., Litjens, G., Hermsen, M., Karssemeijer, N., & van der Laak, J. A. (2015a, March). A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images. In *Medical Imaging 2015: Digital Pathology* (Vol. 9420, pp. 99-104). SPIE. https://doi.org/10.1155/2014/789561

Bejnordi, B. E., Litjens, G., Timofeeva, N., Otte-Höller, I., Homeyer, A., Karssemeijer, N., & Van Der Laak, J. A. (2015b). Stain specific standardization of whole-slide histopathological images. *IEEE Transactions on Medical Imaging*, *35*(2), 404-415. https://doi.org/10.1109/TMI.2015.2476509

BenTaieb, A., Li-Chang, H., Huntsman, D., & Hamarneh, G. (2017). A structured latent model for ovarian carcinoma subtyping from histopathology slides. *Medical Image Analysis*, *39*, 194-205. https://doi.org/10.1016/j.media.2017.04.008

Bhargava, R., & Madabhushi, A. (2016). Emerging themes in image informatics and molecular analysis for digital pathology. *Annual Review of Biomedical Engineering*, *18*, 387-412. https://doi.org/10.1146/annurev-bioeng-112415-114722

Bioimaging. (2015). 4th International Symposium in Applied Bioimaging. http://www.bioimaging2015.ineb.up.pt/

Brunye, T. T., Carney, P. A., Allison, K. H., Shapiro, L. G., Weaver, D. L., & Elmore, J. G. (2014). Eye movements as an index of pathologist visual expertise: A pilot study. *PLOS One*, *9*(8), e103447. https://doi.org/10.1371/journal.pone.0103447

Burçak, K. C., Baykan, Ö. K., & Uğuz, H. (2021). A new deep convolutional neural network model for classifying breast cancer histopathological images and the hyperparameter optimisation of the proposed model. *The Journal of Supercomputing*, *77*(1), 973-989. https://doi.org/10.1007/s11227-020-03321-y

Caie, P. D., Turnbull, A. K., Farrington, S. M., Oniscu, A., & Harrison, D. J. (2014). Quantification of tumour budding, lymphatic vessel density and invasion through image analysis in colorectal cancer. *Journal of Translational Medicine*, *12*, 1-12. https://doi.org/10.1186/1479-5876-12-156

Celik, Y., Talo, M., Yildirim, O., Karabatak, M., & Acharya, U. R. (2020). Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognition Letters*, *133*, 232-239. https://doi.org/10.1016/j.patrec.2020.03.011

Chen, H., Qi, X., Yu, L., & Heng, P. A. (2016). DCAN: deep contour-aware networks for accurate gland segmentation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2487-2496. https://doi.org/10.1109/CVPR.2016.273

Cicerone, M. T., & Camp Jr, C. H. (2019). Potential roles for spectroscopic coherent Raman imaging for histopathology and biomedicine. *In Neurophotonics and Biomedical Spectroscopy* 547-570. Elsevier. https://doi.org/10.1016/B978-0-323-48067-3.00021-4

Das, A., Nair, M. S., & Peter, S. D. (2020). Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: A review. *Journal of Digital Imaging*, *33*(5), 1091-1121. https://doi.org/10.1007/s10278-019-00295-z

De Matos, J., Britto Jr, A. D. S., Oliveira, L. E., & Koerich, A. L. (2019). Histopathologic image processing: A review. *ArXiv Preprint ArXiv: 1904.07900*. https://doi.org/10.48550/arXiv.1904.07900

Dietterich, T. G., Lathrop, R. H., & Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, *89*(1-2), 31-71. https://doi.org/10.1016/S0004-3702(96)00034-3

Dimitriou, N., Arandjelović, O., & Caie, P. D. (2019). Deep learning for whole slide image analysis: An overview. *Frontiers in Medicine*, *6*, 264. https://doi.org/10.3389/fmed.2019.00264

Dimopoulos, S., Mayer, C. E., Rudolf, F., & Stelling, J. (2014). Accurate cell segmentation in microscopy images using membrane patterns. *Bioinformatics*, *30*(18), 2644-2651. https://doi.org/10.1093/bioinformatics/btu302

Doyle, S., Hwang, M., Shah, K., Madabhushi, A., Feldman, M., & Tomaszeweski, J. (2007, April). Automated grading of prostate cancer using architectural and textural image features. *In 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, (pp. 1284-1287). IEEE. https://doi.org/10.1109/ISBI.2007.357094

Doyle, S., Madabhushi, A., Feldman, M., & Tomaszeweski, J. (2006). A boosting cascade for automated detection of prostate cancer from digitized histology. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part II 9*, (pp. 504-511). Springer Berlin Heidelberg. https://doi.org/10.1007/11866763_62

Doyle, S., Monaco, J., Feldman, M., Tomaszewski, J., & Madabhushi, A. (2011). An active learning based classification strategy for the minority class problem: Application to histopathology annotation. *BMC Bioinformatics*, *12*, 1-14. https://doi.org/10.1186/1471-2105-12-424

Duenweg, S. R., Bobholz, S. A., Lowman, A. K., Stebbins, M. A., Winiarz, A., Nath, B., ... & LaViolette, P. S. (2023). Whole slide imaging (WSI) scanner differences influence optical and computed properties of digitized prostate cancer histology. *Journal of Pathology Informatics*, *14*, 100321. https://doi.org/10.1016/j.jpi.2023.100321

Dundar, M. M., Badve, S., Bilgin, G., Raykar, V., Jain, R., Sertel, O., & Gurcan, M. N. (2011). Computerized classification of intraductal breast lesions using histopathological images. *IEEE Transactions on Biomedical Engineering*, *58*(7), 1977-1984. https://doi.org/10.1109/TBME.2011.2110648

Gertych, A., Ing, N., Ma, Z., Fuchs, T. J., Salman, S., Mohanty, S., ... & Knudsen, B. S. (2015). Machine learning approaches to analyze histological images of tissues from radical prostatectomies. *Computerized Medical Imaging and Graphics*, *46*, 197-208. https://doi.org/10.1016/j.compmedimag.2015.08.002

Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, *2*, 147-171. https://doi.org/10.1109/RBME.2009.2034865

Gutman, D., Codella, N. C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., & Halpern, A. (2016). Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). *ArXiv Preprint ArXiv: 1605. 01397*. https://doi.org/10.48550/arXiv.1605.01397

Hameed, Z., Zahia, S., Garcia-Zapirain, B., Javier Aguirre, J., & Maria Vanegas, A. (2020). Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors*, *20*(16), 4373. https://doi.org/10.3390/s20164373

Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., & Li, S. (2017). Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports*, *7*(1), 4172. https://doi.org/10.1038/s41598-017-04075-z

Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E., & Saltz, J. H. (2016). Patch-based convolutional neural network for whole slide tissue image classification. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2424-2433. https://doi.org/10.1109/CVPR.2016.266

Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., & Sun, Q. (2018). Deep learning for image-based cancer detection and diagnosis-A survey. *Pattern Recognition*, *83*, 134-149. http://doi.org/10.1016/j.patcog.2018.05.014

Huang, C. H., & Racoceanu, D. (2018). eXclusive Autoencoder (XAE) for Nucleus Detection and Classification on Hematoxylin and Eosin (H and E) Stained Histopathological Images. *ArXiv Preprint ArXiv: 1811.11243*. https://doi.org/10.48550/arXiv.1811.11243

Jia, Z., Huang, X., Eric, I., Chang, C., & Xu, Y. (2017). Constrained deep weak supervision for histopathology image segmentation. *IEEE Transactions on Medical Imaging*, *36*(11), 2376-2388. https://doi.org/10.1109/TMI.2017.2724070

Jin, X., Xu, A., Bie, R., & Guo, P. (2006). Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. *In Data Mining for Biomedical Applications: PAKDD 2006 Workshop, BioDM 2006, Singapore, April 9, 2006. Proceedings* 106-115. Springer Berlin Heidelberg. https://doi.org/10.1007/11691730_11

Kather, J. N., Weis, C. A., Bianconi, F., Melchers, S. M., Schad, L. R., Gaiser, T., ... & Zöllner, F. G. (2016). Multi-class texture analysis in colorectal cancer histology. *Scientific Reports*, *6*(1), 1-11. https://doi.org/10.1038/srep27988

Kazmar, T., Šmid, M., Fuchs, M., Luber, B., & Mattes, J. (2010, August). Learning cellular texture features in microscopic cancer cell images for automated cell-detection. *In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 49-52. IEEE. https://doi.org/10.1109/IEMBS.2010.5626299

Kieffer, B., Babaie, M., Kalra, S., & Tizhoosh, H. R. (2017, November). Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. *In 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, (pp. 1-6). IEEE. https://doi.org/10.1109/IPTA.2017.8310149

Kohl, M., Walz, C., Ludwig, F., Braunewell, S., & Baust, M. (2018). Assessment of breast cancer histology using densely connected convolutional networks. In *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27-29, 2018, Proceedings 15*, (pp. 903-913). Springer International Publishing. https://doi.org/10.1007/978-3-319-93000-8_103

Kolarević, D., Vujasinović, T., Kanjer, K., Milovanović, J., Todorović-Raković, N., Nikolić-Vukosavljević, D., & Radulovic, M. (2018). Effects of different preprocessing algorithms on the prognostic value of breast tumour microscopic images. *Journal of Microscopy*, *270*(1), 17-26. https://doi.org/10.1111/jmi.12645

Kothari, S., Phan, J. H., & Wang, M. D. (2013). Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade. *Journal of Pathology Informatics*, *4*(1), 22. https://doi.org/10.4103/2153-3539.117448

Krithiga, R., & Geetha, P. (2021). Breast cancer detection, segmentation and classification on histopathology images analysis: A systematic review. *Archives of Computational Methods in Engineering*, *28*(4), 2607-2619. https://doi.org/10.1007/s11831-020-09470-w

Lafarge, M. W., Pluim, J. P., Eppenhof, K. A., Moeskops, P., & Veta, M. (2017). Domain-adversarial neural networks to address the appearance variability of histopathology images. *In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 3rd International Workshop, DLMIA 2017 and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, (pp. 83-91). Springer International Publishing. https://doi.org/10.1007/978-3-319-67558-9_10

Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., ... & Khalil, T. (2019). Content-based image retrieval and feature extraction: A comprehensive review. *Mathematical Problems in Engineering*, *2019*. https://doi.org/10.1155/2019/9658350

Lee, J., Yoo, S. K., Kim, K., Lee, B. M., Park, V. Y., Kim, J. S., & Kim, Y. B. (2023). Machine learning-based radiomics models for prediction of locoregional recurrence in patients with breast cancer. *Oncology Letters*, *26*(4), 1-10. https://doi.org/10.3892/ol.2023.14008

Lee, S., Farley, C., Shim, S., Yoo, W. S., Zhao, Y., & Choi, W. (2020, October). Unsupervised Learning of Deep-Learned Features from Breast Cancer Images. *In 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, (pp. 740-745). IEEE. https://doi.org/10.1109/BIBE50027.2020.00126

Li, X., & Plataniotis, K. N. (2015, March). Color model comparative analysis for breast cancer diagnosis using H and E stained images. *In Medical Imaging 2015: Digital Pathology* (Vol. *9420*, pp. 137-142). SPIE. https://doi.org/10.1117/12.2079935

Lin, H., Chen, H., Dou, Q., Wang, L., Qin, J., & Heng, P. A. (2018, March). Scannet: A fast and dense scanning framework for metastastic breast cancer detection from whole-slide image. *In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, (pp. 539-546). IEEE. https://doi.org/10.1109/WACV.2018.00065

Lin, T. Y., RoyChowdhury, A., & Maji, S. (2015). Bilinear CNN models for fine-grained visual recognition. *In Proceedings of the IEEE International Conference on Computer Vision*, (pp. 1449-1457). https://www.cv-foundation.org/openaccess/content_iccv_2015/html/Lin_Bilinear_CNN_Models_ICCV_2015_paper.html

Linder, N., Konsti, J., Turkki, R., Rahtu, E., Lundin, M., Nordling, S., ... & Lundin, J. (2012). Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagnostic Pathology*, *7*, 1-11. https://doi.org/10.1186/1746-1596-7-22

Litjens, G., Bandi, P., Ehteshami Bejnordi, B., Geessink, O., Balkenhol, M., Bult, P., ... & van der Laak, J. (2018). 1399 H and E-stained sentinel lymph node sections of breast cancer patients: The CAMELYON dataset. *GigaScience*, *7*(6), giy065. https://doi.org/10.1093/gigascience/giy065

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60-88. https://doi.org/10.1016/j.media.2017.07.005

Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., ... & Van Der Laak, J. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports*, *6*(1), 26286. https://doi.org/10.1038/srep26286

Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., Kohlberger, T., Boyko, A., ... & Stumpe, M. C. (2017). Detecting cancer metastases on gigapixel pathology images. *ArXiv Preprint ArXiv:1703.02442*. https://doi.org/10.48550/arXiv.1703.02442

Ludovic, R., Daniel, R., Nicolas, L., Maria, K., Humayun, I., Jacques, K., ... & Catherine, G. (2013). Mitosis detection in breast cancer histological images an ICPR 2012 contest. *Journal of Pathology Informatics*, *4*(1), 8. https://doi.org/10.4103/2153-3539.112693

Medjahed, S. A., Saadi, T. A., & Benyettou, A. (2013). Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. *International Journal of Computer Applications*, *62*(1). https://doi.org/10.5120/10041-4635

Mercan, E., Aksoy, S., Shapiro, L. G., Weaver, D. L., Brunyé, T. T., & Elmore, J. G. (2016). Localization of diagnostically relevant regions of interest in whole slide images: A comparative study. *Journal of Digital Imaging*, *29*, 496-506. https://doi.org/10.1007/s10278-016-9873-1

Miyato, T., Maeda, S. I., Koyama, M., & Ishii, S. (2018). Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(8), 1979-1993. https://doi.org/10.1109/TPAMI.2018.2858821

Mungle, T., Tewary, S., Das, D. K., Arun, I., Basak, B., Agarwal, S., ... & Chakraborty, C. (2017). MRF-ANN: a machine learning approach for automated ER scoring of breast cancer immunohistochemical images. *Journal of Microscopy*, *267*(2), 117-129. https://doi.org/10.1111/jmi.12552

Nalisnik, M., Amgad, M., Lee, S., Halani, S. H., Velazquez Vega, J. E., Brat, D. J., ... & Cooper, L. A. (2017). Interactive phenotyping of large-scale histology imaging data with HistomicsML. *Scientific Reports*, *7*(1), 14588. c10.1038/s41598-017-15092-3

Niwas, S. I., Palanisamy, P., & Sujathan, K. (2010, July). Complex wavelet based texture features of cancer cytology images. In *2010 5th International Conference on Industrial and Information Systems*, 348-353. IEEE. https://doi.org/10.1109/ICIINFS.2010.5578679

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, *29*(1), 51-59. https://doi.org/10.1016/0031-3203(95)00067-4

Padmanabhan, R. K., Somasundar, V. H., Griffith, S. D., Zhu, J., Samoyedny, D., Tan, K. S., ... & Lee, W. M. (2014). An active learning approach for rapid characterization of endothelial cells in human tumors. *PLOS One*, *9*(3), e90495. https://doi.org/10.1371/journal.pone.0090495

Pan, Y., Birdwell, J. D., & Djouadi, S. M. (2006, October). An efficient bottom-up image segmentation method based on region growing, region competition and the mumford shah functional. *In 2006 IEEE Workshop on Multimedia Signal Processing*, (pp. 344-349). IEEE. https://doi.org/10.1109/MMSP.2006.285327

Pankaja, L. P. & Sivagami M. (2024). LT-LBP-Based Spatial Texture Feature Extraction with Deep Learning for X-Ray Images. *Journal of Computer Science*, *20*(1), 106-120. https://doi.org/10.3844/jcssp.2024.106.120

Peikari, M., Zubovits, J., Clarke, G., & Martel, A. L. (2015). Clustering analysis for semi-supervised learning improves classification performance of digital pathology. In *Machine Learning in Medical Imaging: 6th International Workshop, MLMI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Proceedings 6*, (pp. 263-270). Springer International Publishing. https://doi.org/10.1007/978-3-319-24888-2_32

Ponraj, A., & Canessane, A. (2023). Radial basis function networks and contrast-limited adaptive histogram equalization filter based early-stage breast cancer detection techniques. *J Comput Sci*, *19*, 760-74. https://doi.org/10.3844/jcssp.2023.760.774

Prvulović, I., Kardum-Skelin, I., Susterčić, D., Jakić-Razumović, J., & Manojlović, S. (2010). Morphometry of tumor cells in different grades and types of breast cancer. *Collegium Antropologicum*, *34*(1), 99-103. https://hrcak.srce.hr/51249

Raghunath, V., Braxton, M. O., Gagnon, S. A., Brunyé, T. T., Allison, K. H., Reisch, L. M., ... & Shapiro, L. G. (2012). Mouse cursor movement and eye tracking data as an indicator of pathologists' attention when viewing digital whole slide images. *Journal of Pathology Informatics*, *3*(1), 43. https://doi.org/10.4103/2153-3539.104905

Rasmus, A., Berglund, M., Honkala, M., Valpola, H., & Raiko, T. (2015). Semi-supervised learning with ladder networks. *Advances in Neural Information Processing Systems*, *28*. https://doi.org/10.48550/arXiv.1507.02672

Rexhepaj, E., Agnarsdóttir, M., Bergman, J., Edqvist, P. H., Bergqvist, M., Uhlén, M., ... & Ponten, F. (2013). A texture based pattern recognition approach to distinguish melanoma from non-melanoma cells in histopathological tissue microarray sections. *PLOS One*, *8*(5), e62070. https://doi.org/10.1371/journal.pone.0062070

Romo, D., García-Arteaga, J. D., Arbeláez, P., & Romero, E. (2014, March). A discriminant multi-scale histopathology descriptor using dictionary learning. *In Medical Imaging 2014: Digital Pathology* (Vol. *9041*, pp. 167-172). SPIE. https://doi.org/10.1117/12.2043935

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*, 211-252. https://doi.org/10.1007/s11263-015-0816-y

Saha, M., Chakraborty, C., Arun, I., Ahmed, R., & Chatterjee, S. (2017). An advanced deep learning approach for Ki-67 stained hotspot detection and proliferation rate scoring for prognostic evaluation of breast cancer. *Scientific Reports*, *7*(1), 3213. https://doi.org/10.1038/s41598-017-03405-5

Saito, A., Numata, Y., Hamada, T., Horisawa, T., Cosatto, E., Graf, H. P., ... & Yamamoto, Y. (2016). A novel method for morphological pleomorphism and heterogeneity quantitative measurement: Named cell feature level co-occurrence matrix. *Journal of Pathology Informatics*, *7*(1), 36. https://doi.org/10.4103/2153-3539.189699

Salman, S., Ma, Z., Mohanty, S., Bhele, S., Chu, Y. T., Knudsen, B., & Gertych, A. (2014). A machine learning approach to identify prostate cancer areas in complex histological images. In *Information Technologies in Biomedicine, Volume 3*, 295-306. Springer International Publishing. https://doi.org/10.1007/978-3-319-06593-9_26

Sathish, D., Kamath, S., Prasad, K., Kadavigere, R., & Martis, R. J. (2017). Asymmetry analysis of breast thermograms using automated segmentation and texture features. *Signal, Image and Video Processing*, *11*, 745-752. https://doi.org/10.1007/s11760-016-1018-y

Selvi, C., & Suganthi, M. (2018). A novel enhanced gray scale adaptive method for prediction of breast cancer. *Journal of Medical Systems*, *42*(11), 221. https://doi.org/10.1007/s10916-018-1082-7

Shah, M., Wang, D., Rubadue, C., Suster, D., & Beck, A. (2017, November). Deep learning assessment of tumor proliferation in breast cancer histological images. *In 2017 IEEE International Conference on Bioinformatics and Biomedicine* (*BIBM*), 600-603. IEEE. https://doi.org/10.1109/BIBM.2017.8217719

Sharma, S., & Mehra, R. (2020). Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images-a comparative insight. *Journal of Digital Imaging*, *33*(3), 632-654. https://doi.org/10.1007/s10278-019-00307-y

Sheikhzadeh, F., Ward, R. K., van Niekerk, D., & Guillaud, M. (2018). Automatic labeling of molecular biomarkers of immunohistochemistry images using fully convolutional networks. *PLOS One*, *13*(1), e0190783. https://doi.org/10.1371/journal.pone.0190783

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, *19*, 221-248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285-1298. https://doi.org/10.1109/TMI.2016.2528162

Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P. A., Guo, Y. B., ... & Rajpoot, N. M. (2017). Gland segmentation in colon histology images: The glas challenge contest. *Medical Image Analysis*, *35*, 489-502. https://doi.org/10.1016/j.media.2016.08.008

Song, Y., Zhang, L., Chen, S., Ni, D., Lei, B., & Wang, T. (2015). Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. *IEEE Transactions on Biomedical Engineering*, *62*(10), 2421-2433. https://doi.org/10.1109/TBME.2015.2430895

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016, July). Breast cancer Convolutional neural netwoon using convolutional neural networks. *In 2016 International Joint Conference on Neural Networks (IJCNN)*, 2560-2567. IEEE. https://doi.org/10.1109/IJCNN.2016.7727519

Sparks, R., & Madabhushi, A. (2016). Out-of-sample extrapolation utilizing semi-supervised manifold learning (OSE-SSL): Content based image retrieval for histopathology images. *Scientific Reports*, *6*(1), 27306. https://doi.org/10.1038/srep27306

Sridhar, A., Doyle, S., & Madabhushi, A. (2015). Content-based image retrieval of digitized histopathology in boosted spectrally embedded spaces. *Journal of Pathology Informatics*, *6*(1), 41. https://doi.org/10.4103/2153-3539.159441

Srinidhi, C. L., Ciga, O., & Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, *67*, 101813. https://doi.org/10.1016/j.media.2020.101813

Sukumar, P., & Gnanamurthy, R. K. (2016). Computer aided detection of cervical cancer using pap smear images based on adaptive neuro fuzzy inference system classifier. *Journal of Medical Imaging and Health Informatics*, *6*(2), 312-319. https://doi.org/10.1166/jmihi.2016.1690

Tosta, T. A. A., Neves, L. A., & do Nascimento, M. Z. (2017). Segmentation methods of H and E-stained histological images of lymphoma: A review. *Informatics in Medicine Unlocked*, *9*, 35-43. https://doi.org/10.1016/j.imu.2017.05.009

Vang, Y. S., Chen, Z., & Xie, X. (2018). Deep learning framework for multi-class breast cancer histology image classification. In *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27-29, 2018, Proceedings 15*, (pp. 914-922). Springer International Publishing. https://doi.org/10.1007/978-3-319-93000-8_104

Vesal, S., Ravikumar, N., Davari, A., Ellmann, S., & Maier, A. (2018). Classification of breast cancer histology images using transfer learning. In *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27-29, 2018, Proceedings 15* 812-819. Springer International Publishing. https://doi.org/10.1007/978-3-319-93000-8_92

Veta, M., Pluim, J. P., Van Diest, P. J., & Viergever, M. A. (2014). Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*, *61*(5), 1400-1411. https://doi.org/10.1109/TBME.2014.2303852

Vink, J. P., Van Leeuwen, M. B., Van Deurzen, C. H. M., & de Haan, G. (2013). Efficient nucleus detector in histopathology images. *Journal of Microscopy*, *249*(2), 124-135. https://doi.org/10.1111/jmi.12001

Wadhwa, G., & Kaur, A. (2020, July). A deep cnn technique for detection of breast cancer using histopathology images. In *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)* (pp. 179-185). IEEE. https://doi.org/10.1109/ACCTHPA49271.2020.9213192

Wang, C., Shi, J., Zhang, Q., & Ying, S. (2017, July). Histopathological image classification with bilinear convolutional neural networks. *In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4050-4053. IEEE. https://doi.org/10.1109/EMBC.2017.8037745

Wang, D., Foran, D. J., Ren, J., Zhong, H., Kim, I. Y., & Qi, X. (2015, August). Exploring automatic prostate histopathology image gleason grading via local structure modeling. *In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2649-2652. IEEE. https://doi.org/10.1109/EMBC.2015.7318936

Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016a). Deep learning for identifying metastatic breast cancer. *ArXiv Preprint ArXiv: 1606.05718*. https://doi.org/10.48550/arXiv.1606.05718

Wang, P., Hu, X., Li, Y., Liu, Q., & Zhu, X. (2016b). Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Processing*, *122*, 1-13. https://doi.org/10.1016/j.sigpro.2015.11.011

Wang, S., Burtt, K., Turkbey, B., Choyke, P., & Summers, R. M. (2014). Computer aided-diagnosis of prostate cancer on multiparametric MRI: A technical review of current research. *BioMed Research International*, *2014*. https://doi.org/10.1155/2014/789561

Wang, Y., Williamson, K. E., Kelly, P. J., James, J. A., & Hamilton, P. W. (2012). SurfaceSlide: A multitouch digital pathology platform. *PLOS One*, *7*(1), e30783. https://doi.org/10.1371/journal.pone.0030783

WCRFI. (2020). Global cancer incidence in women. *American Institute for Cancer Research*. https://www.wcrf.org/cancer-trends/worldwide-cancer-data

WHO. (2022). Cancer. *World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/cancer

Wu, H., Phan, J. H., Bhatia, A. K., Cundiff, C. A., Shehata, B. M., & Wang, M. D. (2015, August). Detection of blur artifacts in histopathological whole-slide images of endomyocardial biopsies. *In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (pp. 727-730). IEEE. https://doi.org/10.1109/EMBC.2015.7318465

Xing, F., & Yang, L. (2016). Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. *IEEE Reviews in Biomedical Engineering*, *9*, 234-263. https://doi.org/10.1109/RBME.2016.2515127

Xu, J., Luo, X., Wang, G., Gilmore, H., & Madabhushi, A. (2016). A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing*, *191*, 214-223. https://doi.org/10.1016/j.neucom.2016.01.034

Xu, Y., Li, Y., Shen, Z., Wu, Z., Gao, T., Fan, Y., ... & Chang, E. I. C. (2017a). Parallel multiple instance learning for extremely large histopathology image analysis. *BMC Bioinformatics*, *18*, 1-15. https://doi.org/10.1186/s12859-017-1768-8

Xu, Y., Jia, Z., Wang, L. B., Ai, Y., Zhang, F., Lai, M., & Chang, E. I. C. (2017b). Large scale tissue histopathology image classification, segmentation and visualization via deep convolutional activation features. *BMC Bioinformatics*, *18*, 1-17. https://doi.org/10.1186/s12859-017-1685-x

Xu, Y., Zhu, J. Y., Eric, I., Chang, C., Lai, M., & Tu, Z. (2014). Weakly supervised histopathology cancer image segmentation and classification. *Medical Image Analysis*, *18*(3), 591-604. https://doi.org/10.1016/j.media.2014.01.010

Yan, R., Ren, F., Wang, Z., Wang, L., Zhang, T., Liu, Y., ... & Zhang, F. (2020). Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, *173*, 52-60. https://doi.org/10.1016/j.ymeth.2019.06.014

Yari, Y., Nguyen, T. V., & Nguyen, H. T. (2020). Deep learning applied for histological diagnosis of breast cancer. *IEEE Access*, *8*, 162432-162448. https://doi.org/10.1109/ACCESS.2020.3021557

Yuan, B., Schafferer, S., Tang, Q., Scheffler, M., Nees, J., Heil, J., ... & Burwinkel, B. (2019). A plasma metabolite panel as biomarkers for early primary breast cancer detection. *International Journal of Cancer*, *144*(11), 2833-2842. https://doi.org/10.1002/ijc.31996

Zainudin, Z., Shamsuddin, S. M., & Hasan, S. (2020). Deep layer CNN architecture for breast cancer histopathology image detection. In *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019) 4*, (pp. 43-51). Springer International Publishing. https://doi.org/10.1007/978-3-030-14118-9_5

Zalloum, H. N. Zeer, S. A. Manassra, A. Sara, M. R. A. & Alkhateeb, J. H. (2022). Breast cancer grading using machine learning approach algorithms. *Journal of Computer Science*, *18*(12), 1213-1218. https://doi.org/10.3844/jcssp.2022.1213.1218

Zhang, W., Itoh, K., Tanida, J., & Ichioka, Y. (1990). Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Applied Optics*, *29*(32), 4790-4797. https://doi.org/10.1364/AO.29.004790

Zhu, Y., Zhang, S., Liu, W., & Metaxas, D. N. (2014). Scalable histopathological image analysis via active learning. *In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part III 17*, (pp. 369-376). Springer International Publishing. https://doi.org/10.1007/978-3-319-10443-0_47

Žunić, D., & Žunić, J. (2014). Shape ellipticity from Hu moment invariants. *Applied Mathematics and Computation*, *226*, 406-414. https://doi.org/10.1016/j.amc.2013.10.062