CHAPTER 1

Analysis of Histopathology Images:
From Traditional Machine Learning to Deep Learning

Oscar Jimenez-del-Toro[*,**], Sebastian Otálora[*], Mats Andersson[†], Kristian Eurén[†], Martin Hedlund[†], Mikael Rousson[†], Henning Müller[*,**] and Manfredo Atzori[*]

∗ University of Applied Sciences Western Switzerland (HES–SO), Information Systems Institute, rue du TechnoPole 3, 3960 Sierre Switzerland Corresponding author: oscar.jimenez@hevs.ch
∗∗ Unversity of Geneva, Department of Radiology and Medical Informatics, Geneva, Switzerland
† ContextVision, Stockholm, Sweden

Abstract

Digitizing pathology is a current trend that makes large amounts of visual data available for automatic analysis. It allows to visualize and interpret pathologic cell and tissue samples in high–resolution images and with the help of computer tools. This opens the possibility to develop image analysis methods that help pathologists and support their image descriptions (i.e. staging, grading) with objective quantification of image features. Numerous detection, classification and segmentation algorithms of the underlying tissue primitives in histopathology images have been proposed in this respect. To better select the most suitable algorithms for histopathology tasks, biomedical image analysis challenges have evaluated and compared both traditional feature extraction with machine learning and deep learning techniques. This chapter provides an overview of methods addressing the analysis of histopathology images, as well as a brief description of the tasks they aim to solve. It is focused on histopathology images containing textured areas of different types.

Keywords: Histopathology, Deep learning, Biomedical texture analysis, Digital pathology

1. Histopathology Imaging: a Challenge for Texture Analysis

Histopathology is the examination of a biopsy or surgical tissue specimen by a pathologist. The image analysis of histopathological samples is used to provide the final detailed diagnosis of several diseases, including most cancers. Current histopathology practice has several constraints, as it is highly time–consuming and often shows low agreement between pathologists. In this context, Computer–Assisted Diagnosis (CAD) of histopathology images is a novel challenging field for biomedical image analysis. CAD of histopathology images can help to solve these constraints, because histopathology images are characterized by repetitive patterns at several scales that

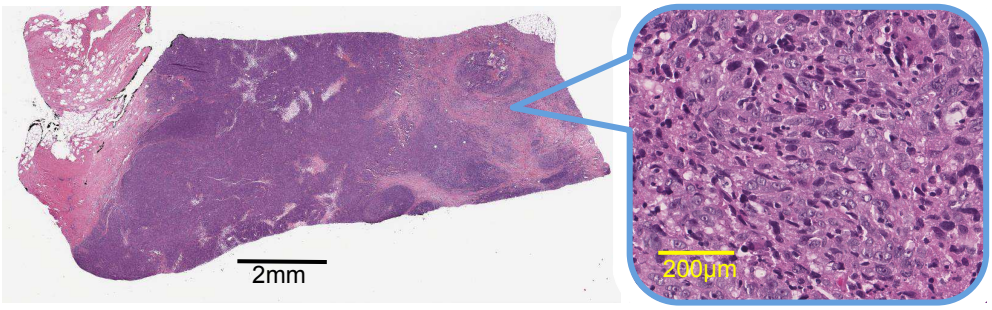


Figure 1.1 Example of a histopathology whole slide image, to the right, and a high power field (image zoomed at a high resolution) to the left.

can be particularly suited for texture analysis and automated recognition. The fusion of traditional diagnosis methods with computational data analysis represents an opportunity to reduce the workload of pathologists while also harmonizing performance.

The standard process of histopathology image preparation passes through several phases that highlight specific structures in the images. First, the tissue is fixated and put onto glass slides with chemicals or by freezing the sections. Then, the sections can be stained with pigments (e.g. haematoxylin–eosin, H&E), using antibodies (immunohistochemistry, IHC) or with other methods (e.g. immuno–fluorescence labeling). Afterwards, the tissue slides are examined under a microscope by the pathologist.The pathologist's visual inspection of tissue samples is currently the standard method to diagnose a considerable number of diseases and to grade/stage most types of cancer [52, 97]). When CAD is considered, there are two additional phases in the work flow. First, the image is scanned (generally using whole slide imaging). Then, the digital images are processed and analyzed using computer–based methods such as visual feature extraction and machine learning.

The clinical analysis of histopathology images can be laborious for pathologists and prone to inconsistencies. The staging and grading of cancer is getting increasingly complex due to cancer incidence and patient–specific treatment options. The detailed analysis of a single case could require several slides with multiple stainings. Moreover, quantitative parameters are increasingly required (such as mitoses counting) [77]. Specific protocols were created to analyze biopsies or resected tissue specimens for the most common cancer types (such as lung, breast or prostate). These protocols have led to precise and widely accepted prognostic grading strategies, for example the Gleason grading system for prostate cancer [47]. However, the diagnostic practice is increasing the pressure for pathologists to handle large volumes of cases while providing a larger amount of information in the pathology reports [44]. Pathol-

ogists now spend much time on benign biopsies that represent approximately 80% of all biopsies [52]. The inter–observer agreement of pathologists for grading the same slide can be low with some cases reported between 0.52 and 0.73 and between 0.35 and 0.65 [40]. A second inspection of the same specimens can lead to 2.3% of major disagreements [82]. This disagreement between pathologists can lead to substantial changes in patient treatment [121, 82].

Digital pathology has its origins in the 1980s, but many factors prevented it from being used in clinical practice including slow scanning speed, poor quality on screen, high costs, required memory and limited network bandwidth. The first high–resolution, automated, whole–slide imaging (WSI) system was developed in 1999 [55]. Since then, the interest in WSI for pathology has continuously grown [120]. Whole slide imaging has been a strong advance for pathology because it overcomes the limitations of previous image acquisition methods, such as poor image quality and image navigation [120]. After whole slide imaging, the shift to a fully digital environment is currently expected for pathology, just as it previously happened for radiology. Digital pathology holds tremendous opportunities for histopathology practice: (1) it can allow online consultations; (2) it can provide access to pathology services in remote locations with limited pathology support; (3) it allows improving the productivity of pathologists by easily accessing and searching digital image archives (an activity representing up to 15% of the pathologists' work) [108]. Nevertheless, the advantages of digital pathology are not yet sufficient to overcome its limitations: WSI requires considerably large storage volumes (each WSI can require 2-3 Gb) [59], scanning is an additional step in the prediagnostic work flow (thus representing an increase in the workload of pathologists or lab technicians) and the quality needs to be certified and accepted by pathologists. The number of scientific researchers and companies developing CAD algorithms for pathology has strongly increased in the last 10 years, further indicating that pathology is becoming digital. Computer–assisted diagnosis of WSI can help overcome the current limitations in digital pathology, promoting its use while reducing the workload of pathologists and rater disagreement.

Texture is generally characterized by homogenous areas with properties related to scale and regular patterns that can occur in 2D or 3D (see Chapter. CROSS-REF_Fundamentals of Texture Processing for Biomedical Image Analysis section Biomedical texture processes) [31]. Biomedical tissue samples contain a complex mixture of repetitive visual patterns at different scales, revealing that organ systems are composed of a few groups of tissue types (e.g. connective, epithelial, muscle and nervous tissue) built from specific cells. The acquisition process (cutting bidimensional sections out of three dimensional structures), results in a partial representation of the organs and systems. It allows the visualization of alterations in the morphology of the tissue architecture that are associated with types and grades of disease. Several approaches have been presented to provide CAD of histopathology images but the field still faces

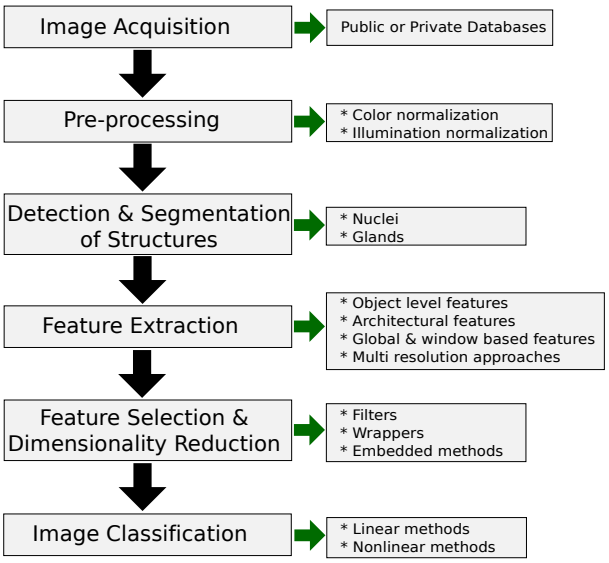


Figure 1.2 The traditional machine learning approach to histopathology CAD.

several challenges. CAD methods can be subdivided into traditional machine learning approaches (e.g. based on the detection of specific structures, texture analysis) and deep learning approaches, as will be discussed in Section 2 and 3, respectively. The current challenges are presented with concrete applications, such as the analysis of specific pathologies, the detection of mitosis, the classification of regions or WSI and the segmentation of specific structures, providing a description of the most recent trends and achievements.

2. Traditional Machine Learning Approaches

Traditional machine learning approaches often include several phases to deal with histopathology images, as represented in Figure 1.2. Each phase is described in the following sections.

2.1. Pre–processing

Pre–processing can compensate for differences between images that are diverse in color, illumination and other defects, such as noise or artifacts that are often due to the scanning process.

2.1.1. Staining normalization

Digital pathology images can have strong color differences due to factors such as the use of different scanners, different stains or staining procedures, section thickness and

sample age. Despite standardization, a perfect color calibration among samples is hard to achieve [79], thus requiring color normalization. Examples of color normalization include histogram–based approaches and color deconvolution–based approaches [69, 80].

2.1.2. Illumination normalization

Shading correction uses empty images (recorded under each considered magnification and illumination condition) to correct the images [83]. Gaussian smoothing can also reveal the intrinsic illumination properties of the image [75]. Background estimation can be performed directly on the specimen [91]. Finally, the image can be modelled as a function of the excitation and emission patterns [16].

2.2. Detection and Segmentation of Structures

The presence as well as the number and the morphological characteristics of specific structures (such as nuclei and glands) are fundamental parameters to evaluate the presence and severity of a pathology such as prostate [47], breast [71] and colorectal [118] cancer.

2.2.1. Nuclei and cells

The nuclei are the central organelles of eukaryotic cells containing most of the cell DNA. Nuclei analysis usually involves detection, segmentation and separation (i.e. dividing overlaps). The identification of seed points in the nuclei is required by most nuclei segmentation and counting methods [61]. Many approaches have been proposed in the literature for nuclei detection, including methods based on Euclidean distance map peaks [28], H–maxima transform [114, 67], Hough transform (detecting seed points for circular–shaped structures, requiring heavy computation) [24], multiscale Laplacian of Gaussian (LoG) filters [2], radial symmetry transform (RST) [109] and (recently) deep learning based methods [5, 23].

Many approaches have also been proposed to perform precise nuclei segmentation. Despite their simplicity, methods based on thresholding and morphological operations can perform well on uniform backgrounds [90, 60] but they are not robust to size, shape and texture changes. The watershed transform requires no tuning but the prior detection of seed points [67, 114]. Active contour models can combine image properties with nuclei shape models [57, 24] but they rely on seed points. Other methods are based on gradients in polar space (GiPS) [28] graph–cuts [105, 2] and machine learning approaches [113]. In H&E stained images, nuclei segmentation accuracy above 90% has been reported [113, 105, 24].

2.2.2. Glands

Glands are organs formed by an ingrowth from an epithelial surface. They synthesize and release substances (such as hormones or mucus) into the bloodstream (endocrine glands) or onto an outer surface of the body, such as the skin or the gastrointestinal tract (exocrine glands). Glands have been segmented automatically, although reliable gland segmentation for diverse cancer grades is still a challenge. Methods based on thresholding and region growing can identify the nuclei and the lumen that are used to initialize seed points for region growing (e.g. [122]). Despite their simplicity, these methods perform well in segmenting healthy and benign glands but they have shown poor results in cancer cases where the gland morphology is deformed. Graph–based methods have been proposed as well (e.g. [4]). Usually, tissue components (such as the nuclei) are identified, represented as vertices and properties of a graph are used to segment the glands. Segmentation based on polar coordinates (the center inside the gland) were proposed on benign and malign glands [43]. Approaches based on Bayesian inference allow to take into account prior knowledge of the structural properties and of the arrangement of glandular components (such as central lumen, surrounding cytoplasm and nuclear periphery) [103, 84].

2.3. Feature Extraction

Histopathology images have been analyzed using many descriptors based on the knowledge of domain experts. Diagnosis criteria are mainly expressed with cytologic terms representing objects (i.e. nuclei, cells, glands) and their presence in malign and benign areas. Thus, several articles approached the problem at the object–level (using segmented object features) and the object relationship level (using architectural features). Multi–resolution global and window–based features have also being used.

2.3.1. Object–level features

Object–level features depend strongly on the considered objects (usually nuclei or glands) and on the segmentation algorithms [52]. These features are relevant for any resolution, but in most cases they are extracted from high–resolution images. Object–level features are usually extracted for each color channel and can be grouped into size and shape features (e.g. area, eccentricity, reflection symmetry), radiometric and densitometric features (e.g. image bands, intensity, hue), texture features (e.g. co–occurrence matrix, run–length and wavelet features) and chromatin–specific features (mean and integrated optical density) [14].

2.3.2. Architectural features

Architectural features are mainly based on graphs (mathematical structures used to model pairwise relations between objects). They are made up of vertices (also called nodes or points) that are connected by edges (arcs or lines). Graphs are an effective

method to represent architectural information using topological features in histopathology images. Graph structures used in histopathology include: the Voronoi tesselation, the Delaunay triangulation, minimum spanning trees, O'Callaghan Neighborhood graphs, connected graphs, relative neighbor graphs, k–NN graphs [52]. Each of these structures allows the computation of several descriptive features used for tissue classification, such as number of nodes, number of edges, edge length, number of triangles, area and eccentricity. A detailed description of graph structures and features can be found in [14]. In the last decade graph–based features have been investigated regularly as they are well suited to characterize tumor architecture. Different graphs can be built, usually to represent the spatial organization of epithelial nuclei. Demir et al. [29] performed a 3 class classification (healthy, malignant, inflammation) in brain cancer biopsies using a complete weighted graph. In Altunbay et al. [4], Delaunay triangulation (DT) graphs included colon tissue components (nuclei, stroma and lumen) as nodes. Chekkoury et al. [18] combined morphologic, network and texton–based features for breast cancer diagnosis. In a fundamental study, Doyle et al. [35] presented a 90% accuracy in distinguishing between cancer and benign tissue for prostate cancer using a combination of features.

2.3.3. Global and window based features

These features can be related to color, texture (for example co–occurrence matrices, run–length and wavelet features). Average object and architecture features can be computed globally or in windows at multiple resolutions. At low resolution, color and texture analysis are often used to indirectly measure the properties of local constituents or of tissue architecture (e.g. density of nuclei and the overall pattern of glands or stroma). General image statistics, histograms, gradients and many other features can be extracted in order to perform measurements of the regions of interest and filters can also be used to extract local features. For instance, prostate cancer has been detected and graded with texture features such as Gabor filters [35, 36, 37], fractal dimension [58], wavelets [62], and morphological operations [33].

2.3.4. Multi–resolution approaches

Multi–resolution approaches were proposed to deal with the multi–resolution characteristics of histopathology images and to mimic the analytic approach from pathologists. The Gaussian pyramid approach was used to represent the images in multiple resolutions [15]. Features can be extracted separately for each level and combined to classify image tiles. Color and texture features are commonly used at low resolutions. Architectural arrangement of glands and nuclei are used at medium scales. Nucleus and gland related features can be discriminatory at high resolutions [36].

2.4. Feature Selection & Dimensionality Reduction

2.4.1. Feature selection

Feature selection is the process of selecting a subset of relevant features for model construction thus reducing training times, simplifying the models (to make interpretation easier) and improving the chances of generalization, avoiding overfitting. There are three main feature selection strategies: filters (e.g. information gain), wrappers (e.g. search guided by accuracy) and embedded methods (in which the features are added or removed from the model depending on prediction errors). Well–known feature selection algorithms are represented by the sequential forward selection and sequential backward selection [92], sequential floating forward search — one of the one of most frequently used methods in pathology image analysis [52] — and sequential floating backward search [92]. Other available methods are genetic algorithms, simulated annealing, boosting [42] and grafting [89].

2.4.2. Dimensionality reduction

Dimensionality reduction has the following aims: first, to reduce storage space and computation time; second, to remove multi–collinearity (thus improving the performance of machine learning models), third, to improve data visualization (reducing to low dimensions such as 2D or 3D). The data transformation may be linear, as in Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) [119, 9, 66]. Several non–linear dimensionality reduction algorithms (such as manifold learning and non–linear embedding) were developed and showed to be useful in histopathology image analysis [35].)

2.5. Classification

Classification methods aim at identifying the category of a new observation among a set of categories on the basis of a labeled training set. Depending on the task, anatomical structure, tissue preparation and features the classification accuracy varies. Random forests with a set of 18 features (including, e.g. histogram related features, autocorrelation, sum average, variance, entropy, contrast) obtained 83% median accuracy in the classification of prostate tissue into 7 classes (Gleason grade 3, 4 and 5; benign stroma; benign hyperplasia; intraepithelial neoplasia and inflammation) [34]. k–Nearest Neighbors (k–NN) with a set of 14 features (intensity, morphological and texture) from segmented nuclei was tested on the classification of hepatocellular carcinoma [57]. Adaboost with a large set of features was used to segment nuclei with various shapes [113] and to detect suspicious areas on digital prostate histopathology [37]. SVM with radial basis function kernel and grid search were used to classify mitotic vs. non–mitotic regions [81], to distinguish between prostate tissue classes [35], to classify colon adenocarcinoma histopathology images vs. benign images [94] and to

classify four subtypes of meningioma [93]. Multiclassifier ensembles [45] represent another solution to increase accuracy. Several classification methods can be integrated to increase accuracy because a fusion of methods can improve results [32]. For instance, in Huang et al. [57], an SVM–based decision graph classifier with feature subset selection on each decision node improved the classification of hepatocellular carcinoma.

The comparison of the methods is difficult, due to heterogeneity of data sets and evaluation metrics. Several studies compared the performance of classification procedures on prostate (e.g. [58].Alexandratou et al. [3] published a comparative study for prostate cancer diagnosis, showing a good performance for cancer detection, 80.8% accuracy for low-high Gleason grade discrimination, and 77.8% for cancer grading. Despite the existence of these studies, there are still only few images available for each pathology and obtaining good annotations of the images is expensive and time consuming. Databases and challenges have been proposed in the past few years (see Section 4).

3. Deep Learning Approaches

Deep learning (DL) methods are currently the most frequently studied and successful type of machine learning algorithms (see Chapter CROSSREF_Deep Learning in Texture Analysis and its Application to Tissue Image Classification). In the last decade, DL outperformed classical machine learning algorithms with hand–crafted features in diverse fields such as computer vision [72], speech recognition [30], natural language processing [48] and also recently in biomedial fields (e.g. functional genomics) [88]. The adoption of DL techniques in the biomedical image analysis community had a positive impact on several tasks, and automatic analysis of histopathology images is no exception [50, 63]. Deep learning approaches in computer vision are based on the composition (*layers*) of non–linear transformations over the raw input pixels (see Chapter CROSSREF_Deep Learning in Texture Analysis and its Application to Tissue Image Classification Section CROSSREF_Introduction to Convolutional Neural Networks). This composition builds increasingly abstract representations that are learned in a hierarchical fashion [73, 49]. An *architecture* is the arrangement and the interconnection of parameters learned by an optimization algorithm for a given DL approach. An example of a commonly used architecture, the convolutional neural network, for histopathology image analysis is depicted in Figure 1.3.

One of the main characteristics of most DL architectures is that the output of the network is only based on the adjustable internal weights. These weights are learned by the network through iterative forward/backward propagation of the training samples and the errors respectively. This is known as the backpropagation algorithm for neural networks [74]. The internal weights are updated using the error between the ground

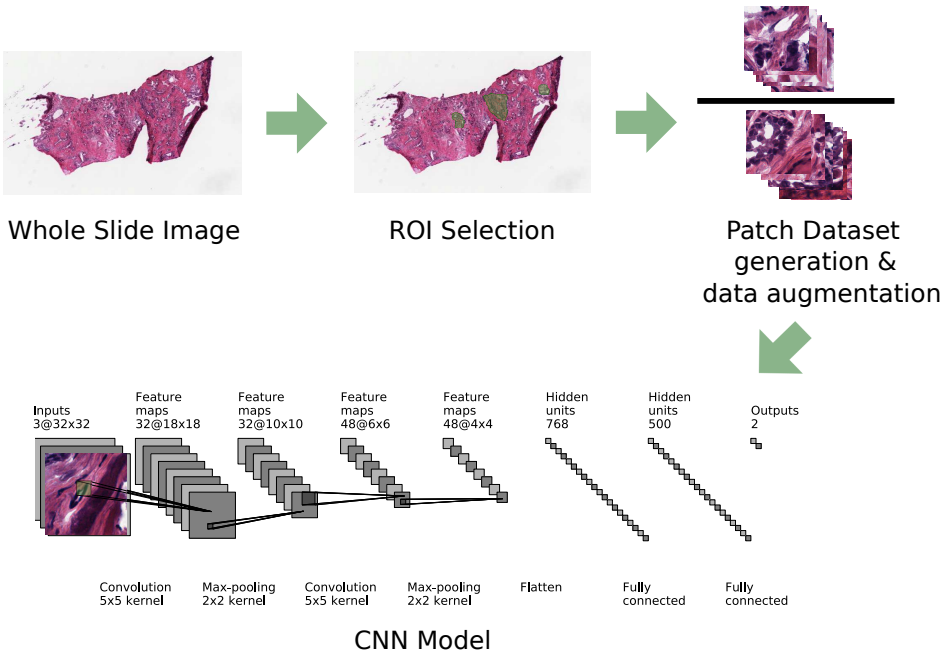


Figure 1.3 A framework for applying supervised deep learning (e.g. CNN) to the classification of a WSI.

truth output of the samples (i.e. the labels in the supervised case and the original representation in the unsupervised one) and the output of the network. This requires less explicit coding of domain knowledge than the traditional machine learning. For a more in depth review of the technical details of the deep learning methods we suggest to the reader to go through the Chapter CROSSREF_Deep Learning in Texture Analysis and its Application to Tissue Image Classification.

There are multiple considerations to take into account when applying a deep learning algorithm to a given histopathology task because the success of the algorithms is partially due to task–specic tuning. Histopathology images are large (i.e. 100.000×100.000 pixels). One of the main characteristics of histopathology images is that the relevant patterns depend on the magnification level. The main considerations are then: the localization of the regions in the image where relevant histopathology primitives are located (e.g. a ROI selection step); the size of the input image (or patch) that is fed to the network and the homogeneity of the staining applied to the WSI (in cases with high variability, staining normalization is required). The architecture of the network plays an important role, although a considerable amount of articles decide to stay with predefined/pretrained network architectures (Figure 1.4). We highlight the main characteristics and risks of DL approaches in histopathology, identifying trends in the

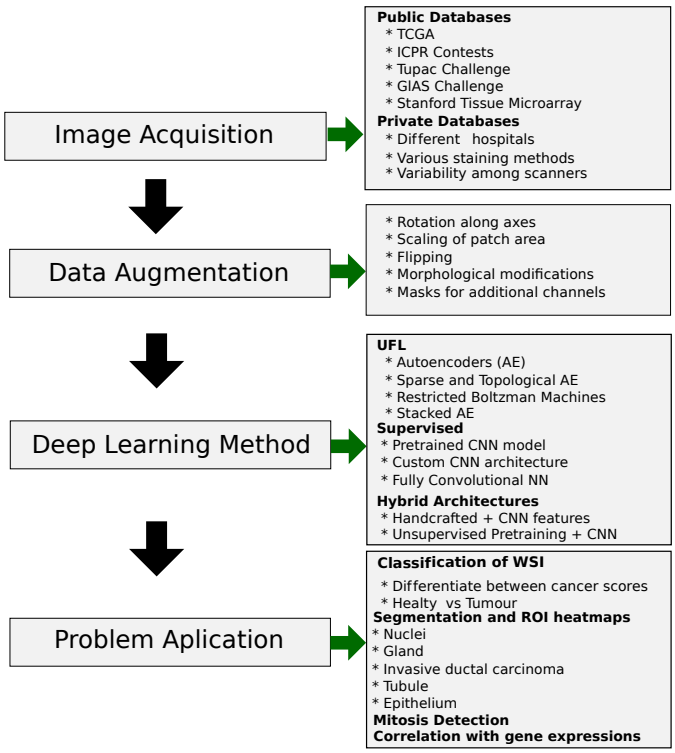


Figure 1.4 Steps involved when applying deep learning to histopathology images.

biological problems and evaluation methods.

3.1. Supervised and Unsupervised Feature Learning Architectures

In machine learning and also in deep learning, there are two families of techniques that differ in supervision. Unsupervised feature learning techniques (UFL) do not depend on labeled samples to detect the inner structure of the data. UFL learns over–complete, sparse or hierarchical unsupervised representations by reconstructing the original input with several constraints (see Figure 1.5 for a basic UFL architecture).

On the other hand, we have supervised techniques that use the label information of the input and guide the model to a desired output. The problem of transitioning between the supervised towards unsupervised learning techniques is one of the biggest challenges for deep learning as discussed by [10, 99]. In unsupervised feature learning, the input data are passed through a compressor to obtain a more compact representation. This is commonly used to subsequently feed a supervised model [10]. UFL representations have been successfully used in the pretraining of deep supervised models [39]. The multiple methods in this family differ in what regularization term is

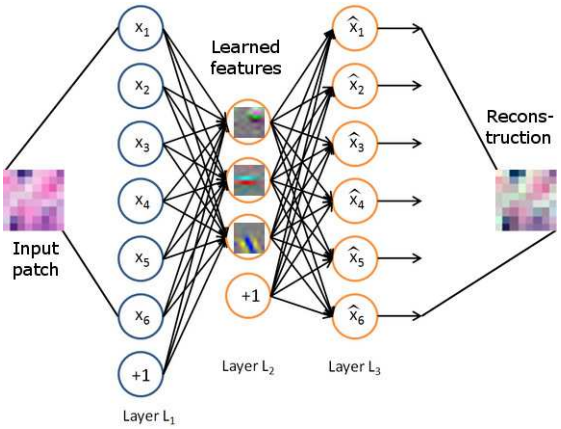


Figure 1.5 A shallow UFL architecture with only one hidden layer that encodes the features to represent a patch.

used.

In formal terms, let $\mathbf{X}$ be the original input (i.e. an RGB patch) then $\mathbf{X} \approx \hat{f}(\mathbf{X}) + r(\theta)$, where $\hat{f}$ is the learned function that reconstructs the input. Usually $\hat{f} = \theta^T\theta$, θ are the parameters of the model, and r is the regularization term that can ensure the sparsity of the learned features or spatial relationships between them.

3.2. Deep Convolutional Neural Networks

CNN models are among the most successful supervised deep learning models for computer vision. The medical imaging field is rapidly adapting these models to solve and improve a large and diverse set of applications [50]. CNNs are a particular kind of a supervised multi layer perceptrons inspired by the visual cortex. The CNN are able to detect visual patterns with minimal pre–processing and are robust in presence of distortion and variability of the pattern. CNNs can usually benefit from data augmentation. Data augmentation consists of subtle transformations in the input data, aimed at learning invariances. The architecture of the CNN is composed of convolutional, pooling and fully connected layers. For a more detailed description of a CNN please refer to Chapter CROSSREF_Deep Learning in Texture Analysis and its Application to Tissue Image Classification Section CROSSREF_Introduction to Convolutional Neural Networks.

3.3. DL Approaches to Histopathology Image Analysis

Most of the deep learning approaches for classification, segmentation and localization in histopathology images are relatively new. The deep neural network methods are not even mentioned in recent reviews of automatic histopathology image analysis, such

as [29, 52, 7]. The first mention of the topic in a review is from Irshad et al. [61]. One of the earliest successful attempts to use deep models in this scenario was from Cireşan et al. [23] at the ICPR 2012 mitosis detection challenge [96]. This work as well as other mitosis detection approaches are described in Section 5 of this Chapter. A more detailed literature review and workflow for deep learning applications in breast cancer is described in Section CROSSREF_Deep Learning Techniques on Texture Analysis of Chest and Breast Images.

The approach described by Cruz–Roa et al. [27] combined unsupervised (UFL) and supervised learning. This method first learns an unsupervised representation via sparse autoencoders and then a convolution and average pooling is performed over this representation to encourage both a translation invariant feature detection and a compact image representation. The proposed convolutional auto–encoder neural network architecture is used for histopathology patch–based image representation learning. This representation allows automatic cancer detection and visually interpretable prediction results analogous to a digital stain. The method identifies image regions that are most relevant for diagnostic decisions, using the probabilities of the final softmax classifier layer of the model. The method was applied to 308 histopathology basal cell carcinoma images at 10X magnification, taking patches of 300×300 pixels. Arevalo et al. [8] computed a hybrid representation for basal–cell carcinoma patches using a topographic unsupervised feature learning method and a bag of features representation. Their approach improved the classification performance by 6% with respect to classical texture based descriptor DCT. In Malon et al. [81], a novel combination of features is proposed. The authors build a feature set of basic statistical measures of the nucleus and cytoplasm pixels and combine them with a CNN classifier. The presented approach improves the performance in comparison with handcrafted features.

Currently, there is a diversification of architectures and applications of deep learning using WSIs. Nayak et al. [85], propose a variation of the unsupervised restricted Boltzmann machine method for learning image signatures. This method classifies patches of clear cell kidney carcinoma and glioblastoma multiforme images from the cancer genome atlas; with this signature the final stage is made using a multi–class regularized support vector classification. Xu et al. [126] developed a technique that deals with few labeled samples. In this work a multiple instance learning framework is introduced, where classification of colon histopathology images is performed. The authors propose a fully supervised approach and a weakly labeled one with similar accuracy (93.56% vs. 94.52%). In the work of Hou et al. [56] a similar idea is proposed. The authors use multiple instance learning to classify between glioblastoma and low–grade glioma images from the cancer genome atlas. The method uses three steps: first it learns masks for discriminative areas using a CNN model with few selected discriminative patches; second, it makes a patch–level prediction using CNNs; third, the class count is performed. In Vanegas et al. [107] the authors do content–based image

retrieval. The proposed technique is based on learning an unsupervised representation based on topographic reconstructed independent component analysis (TICA) and using it as an input for a multimodal semantic indexing method. The authors show that the MAP performance is greater when using the TICA representation in comparison with canonical texture descriptors such as DCT and Haar, showing that UFL based representations also have a good performance in retrieval tasks. In Cruz–Roa et al. [26], the authors propose a 3–layer CNN architecture to detect invasive ductal carcinoma using patches from 113 whole slide images as a training set. This approach obtains 10% higher balanced accuracy in comparison with Haralick and graph–based handcrafted features.

The recent use of unsupervised architectures includes the work of Arevalo et al. [6]. In this work, the authors describe a novel stacked model that shows the best performance when combining features from 2–layered TICA over patches for detecting basal cell carcinoma. The authors also introduce a *digital staining* method based on the weighting of the feature detectors by the classification probability to highlight the areas that are most related to the cancer. Their approach achieved ~0.99 in AUC for a set of 100,000 patches. The approach described by Hang et al. [17] is based on learning a dictionary of 1024 features for classifying kidney clear cell carcinoma and glioblastoma multiforme. The authors used a curated version of the publicly available images from the Cancer Genome Atlas. The dictionary was built with a stacked UFL method called stacked predictive sparse decomposition, which is used in a spatial pyramid matching framework (with the last stage being a linear support vector machine classifier). Noel et al. [86] used a set of 3000 patches extracted from WSIs of the ICPR contests to go a step further in breast cancer detection, classifying each pixel using a CNN into stroma, nuclei, lymphocytes, mitosis, and fat. 90% accuracy was achieved, suggesting that a finer classification of the WSI primitives can help to improve the classification performance.

Besides the public challenges and contests there have been some efforts to do systematic evaluations with the same data sets and experimental setups. Janowczyk and Madabhushi [63] recently published a tutorial on deep learning for digital pathology where they expose exemplar use cases of the technology on more than 1200 slides for 7 tasks including mitosis detection, nuclei segmentation, epithelium segmentation and lymphoma classification. Good results on the studied cases are shown and the source code is provided using the Caffe [64] deep learning framework. Interestingly, the authors did not design a different network architecture for each of the cases but decided to use Alexnet [72] for all of them. Questions such as the importance of manual annotation for ground truth generation are discussed when applying deep learning techniques and also some comparisons against pathologist annotators. It is shown that for some cases the DL system is doing a more refined job on detailed pixel–level something a pathologist can not really do.In Cruz–Roa et al. [25], the authors show an interesting

comparison between unsupervised and supervised approaches for medulloblastoma tumor classification because in this case the proposed CNN architecture is not competitive against UFL methods such as TICA. A further improvement on this dataset was made by Otálora et al. in [87]. The authors combine features from unsupervised and supervised nature (the former from topographic independent component analysis and the latter using Riesz wavelets for two types of medulloblastoma brain cancer images) improving the use of the topographic unsupervised approach. It shows that a combination of a supervised texture representation and UFL contain complementary information about this particular pattern.

Some recent work goes beyond H&E patch classification. A line of work that is closer to the optical properties of the hardware involved in the tissue digitalization is the one presented by Chen et al. [19]. The authors present a new method that aims at detecting flowing colon cancer cells at high–throughput rates, extracting several biophysical features and then building a deep fully connected network. Han et al. [54] present a novel deep UFL method for phenotypic characterization of glioblastoma multiforme. They are able to differentiate two major phenotypic subtypes with different survival curves using the extracted UFL features. Romo–Bucheli et al. [95] quantify tubule nuclei by feeding a CNN model with candidate patches to measure their tubule class membership probability, that was then associated with high–low risk categories determined by an Oncotype DX test. To our knowledge, the application of deep learning stains different from H&E (such as immunohistochemistry) is still not well explored. An exception is the work of Chen et al. [22], in which the authors propose an immune cell detection with a 7–layer CNN model with patches of unmixed colors from the original RGB channels, highlighting immune cell markers. The authors compare their algorithm detection performance with the performance of pathologists achieving up to a 0.99 correlation coefficient. This shows promising results with other types of staining than H&E.

4. Histopathology Challenges

Open scientific challenges targeting tasks in the analysis of pathology images have been proposed in recent years similar to other medical imaging domains[1]. An advantage of having different methods tested on the same data and in the same tasks is the objective comparison of the strengths and limitations of state–of–the–art approaches. Particularly in the case of histopathology, the time–consuming and laborious tasks of searching whole slide images for relevant small tissue primitives (e.g. mitosis, nuclei) can be improved. Selecting the most suitable approach to support and speed–up the visual interpretation of slides can help pathologists to focus on the most important

[1] http://grand-challenge.org/All_Challenges/, *

Table 1.1 Biomedical image analysis challenges related to histopathology.
MD = Mitosis detection, IC = Image classification, SL = Structure localization (segmentation)

Acronym	Challenge	Anatomy	Tasks
2012			
MITOS12 [96]	Mitosis detection in breast cancer histological images	breast	MD
2013			
AMIDA13 [112]	Assessment of mitosis detection algorithms	breast	MD
2014			
MITOS–ATYPIA14	Mitotic count and nuclear pleomorphism	breast	MD,IC
OCCISC [2]	Overlapping cervical cytology image segmentation	cervix	SL
2015			
2OCCISC [3]	2nd Overlapping cervical cytology image segmentation	cervix	SL
GlaS@MICCAI'2015	Gland segmentation challenge	bowel	SL
2016			
SLATMD [53]	Skin lesion analysis towards melanoma detection	skin	SL
CPM	Computational Precision Medicine	lung,neck,brain	IC,SL
CAMELYON16	Cancer metastasis detection in lymph nodes	lymph nodes	IC,SL
TUPAC16	Tumor proliferation assessment challenge	breast	MD,IC

questions when diagnosing these studies. An overview of histopathology related challenges is displayed in Table 1.1. In this section the data sets and tasks from some of these challenges are explained. The scope of this analysis includes only recent digital histopathology challenges, namely MITOS12, AMIDA13, MITOS–ATYPIA-14[4], GlaS@MICCAI'2015[5], CPM[6], CAMELYON16[7] and TUPAC16[8].

4.1. Data sets

The size of the data sets provided to participants for training and testing their algorithms has increased significantly in the latest challenges. In the first challenges, the algorithm's performance was evaluated on regions or high power fields (HPF) from a limited number of WSI (e.g. MITOS12, 50 HPF from 5 slides). However, recent challenges such as CAMELYON16 and TUPAC16 provide a large number (> 500) of WSIs, some of them with manual annotations by pathologists. This promotes the robustness of the participant algorithms as they have to analyze multiple images, sometimes from different scanners and in various disease stages. Table 1.2 shows an overview of the data sets from the digital pathology challenges.

[4] `https://mitos-atypia-14.grand-challenge.org/home/`, *
[5] `http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/glascontest/`, *
[6] `http://miccai.cloudapp.net/competitions/45`, *
[7] `https://camelyon16.grand-challenge.org`, *
[8] `http://tupac.tue-image.nl`, * as of 1 January 2017

Table 1.2 Histopathology challenge data sets. NA = Non applicable NM = Not mentioned

Challenge	Data	Staining	maxResol	Quantity	Scanners
MITOS12	High power fields	H&E	40X	50	2
AMIDA13	High power fields	H&E	40X	23	1
MITOS-ATYPIA14	High power fields	H&E	40X	1136	2
GlaS@MICCAI'2015	High power fields	H&E	20X	165	1
CPM	High power fields	H&E	NP	32	NP
CAMELYON16	Whole slide images	H&E	40X	400	NP
TUPAC16	Whole slide images	H&E	40X	821	1

4.2. Tasks

The complexity of the tasks in the challenges has increased in the past few years. The goals of the challenges can be grouped into three main tasks (Figure 1.6):

- **Mitosis detection**: Identify the mitosis present in high power fields. This has a strong correlation to the aggressiveness of the cancer i.e. faster division of the cells will manifest in more mitosis.
- **Image classification**: Learn the characteristic set of features for a particular tissue class, potentially considering the underlying tissue primitives.
- **Structure localization/segmentation**: Localizing and delineating a boundary around specific tissue structures such as cell nuclei or glands.

Although these tasks are divided into different groups for their description, they can be combined or considered as an initial step for the other tasks, e.g. a whole slide cancer grading algorithm can start by classifying an image as cancer tissue, then segment the tumor and finally grade the WSI based on the mitosis counting inside a tumor region of interest.

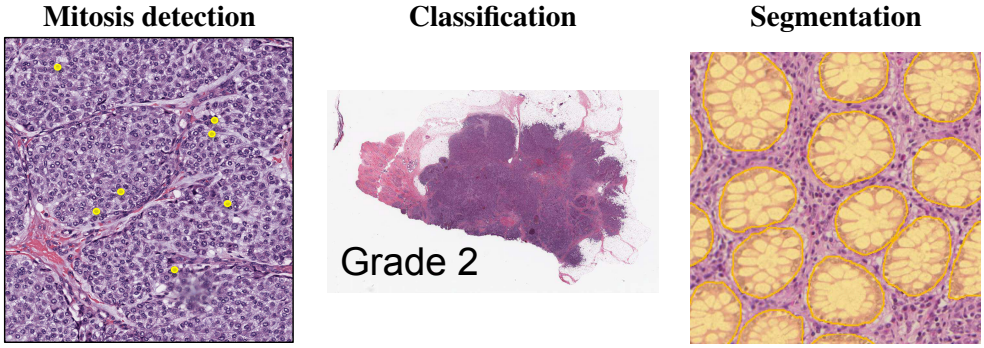


Figure 1.6 Tasks in anatomical pathology challenges.

4.3. Evaluation metrics

An important element of a challenge is the selection of the right evaluation metric. A suitable metric should limit bias in the comparison of the algorithms. It is standard to provide multiple metrics in the challenge results, nevertheless algorithms are generally ranked according to a lead metric for each challenge task. The following metrics were selected for the evaluation of the digital pathology challenges included in Table 1.2.

4.3.1. Mitosis detection

. In the evaluation of mitosis detection algorithms the mitotic events found by the algorithms are counted. Two sets of evaluation metrics have been used in these type of challenges: centroid–based, when each mitotic event is represented by a single location, and region–based when the mitoses are segmented and evaluated as structures with multiple pixels. Still, the centroid–based measures are more frequently used when ranking the participant algorithms in a benchmark. All 4 challenges with a mitosis detection task use the F_1–score as the preferred metric to rank the participant algorithms. The F_1–score is defined as follows:

$$\text{F}_1\text{-score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{1.1}$$

where TP = true positives, FP = false positives, TN = true negatives, FN = false negatives, Precision = TP/(TP + FP) and Recall = TP/(TP + FN). A threshold of $< 7.5 - 8\ \mu m$ (around 30 pixels) was determined as the maximum Euclidean distance from the centroid to consider a mitotic event as a true positive.

4.3.2. Image classification

. Unlike mitosis detection, image classification challenges have used different evaluation approaches for ranking the participant algorithms. The classification can be binary, when the algorithms have to decide if an image contains a tumor/metastasis or not (e.g. CAMELYON16). Otherwise, the classification can be multi–class when the algorithms have to grade the pathologic stage of an image (e.g. TUPAC16). In some cases, the referred challenges also request the probability with which the approaches grade each case (e.g. CAMELYON16) or measure the agreement between the algorithm classification and the pathologist–generated ground truth (e.g. TUPAC16). The proposed evaluation strategies include:

- A points based scheme for nuclear atypia scoring (MITOS–ATYPIA–14).
- Area under a receiver operating characteristic (ROC) curve for the discrimination of lymph node slides containing metastasis or not (CAMELYON16).
- Agreement with ground truth measured with quadratic weighted Cohen's kappa or Spearman's correlation coefficient for breast cancer grading (TUPAC16).

The quadratic weighted Cohen kappa is defined as:

$$K_w = 1 - \frac{\sum_{i,j} w_{i,j} p_{i,j}}{\sum_{i,j} w_{i,j} e_{i,j}} \tag{1.2}$$

where p_{ij}: observed probabilities, $e_{ij} = p_{ij} q_{ij}$: expected probabilities and w_{ij}: weights (with $w_{ij} = w_{ji}$).

4.3.3. Structure localization and/or segmentation

Different type of tissue structures have been targeted for structure segmentation in histopathology challenges: cell nuclei (CPM), metastasis in lymph nodes (CAMELYON16) and colon and rectum glands (GlaS@MICCAI'2015). Three attributes of the output segmentations can be measured: number of structures detected, full structure overlap with the ground truth and shape similarity. The favored segmentation metric (GlaS@MICCAI'2015 and CPM) is the DICE coefficient. Given a ground truth G and S as a set of segmented pixels, DICE is defined as follows:

$$DICE(G, S) = \frac{2|G \cap S|}{|G| + |S|} \tag{1.3}$$

In the GlaS@MICCAI'2015 challenge both the DICE and Hausdorff distance are measured at object level, calculating the score over all the annotated objects included in the test set.

5. Detecting Mitoses

Mitosis detection is an important problem in histopathology image analysis aiming at the identification and annotation of mitotic figures in high power fields. The mitotic figures can then be used as input for a grading system (e.g. the Bloom-Richardson grading system for breast cancer [13]), to have an estimate of the patient's prognosis and to develop personalized treatments. Figure 1.7 shows some examples of mitoses and highlights the difficulty to differentiate between mitotic and non–mitotic cells[9]. Most of the false positive detections arise when apoptotic nuclei (the nuclei going through the natural programmed cell death), are mistaken as mitoses because of their similar appearance. This problem was traditionally addressed with texture descriptors such as DCT, Haar or Haralick [110]. However, traditional techniques are currently struggling to cope with the outstanding performance of deep learning algorithms [112, 86, 20, 100]. Several studies concluded that the performance of DL methods is close to the performance and inter–observer agreement of expert pathologists [111, 22, 46].

[9]as of 1 January of 2017 there is an online tool to compare the performance in mitosis detection: `http://mitosis-detection.herokuapp.com/`

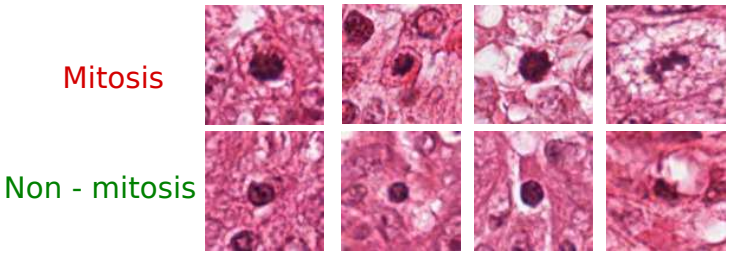


Figure 1.7 Example of hard to differentiate mitotic (top) and non–mitotic (bottom) images.

There have been several challenges on the topic of mitosis detection (Table 1.1). The very first challenge for mitosis detection was the 2012 ICPR *Mitosis Detection in Breast Cancer Histological Images* challenge [96]. 17 teams submitted their results to the contest. The winner was the Swiss IDSIA team [23]. The approach of the team was based on a deep CNN with 13 sequential convolution–maxpooling layers trained over 50 whole slide images with 50 high power fields. This approach was computationally intense because it computed pixel–wise probabilities for mitosis. This method outperformed several classical machine learning classifiers trained with many features describing size, shape, color and texture. Since then, several approaches focused on more clinically usable methods by taking larger areas as input, taking into account only the tumor regions in the WSI and building classifiers for difficult cases (ensemble of architectures among the others) as seen in Figure 1.8.

In Veta et al. [112] the authors summarize the results of the AMIDA13 challenge. The AMIDA13 dataset consisted of 12 training and 11 testing subjects, with more than one thousand annotated mitotic figures by multiple observers. The accuracy of the best method (again, the one by IDSIA) is comparable with the inter–observer variability. IDSIA's method is based on an efficient CNN without any candidate selection, but classifying each of the high power fields pixels.

In 2014, ICPR launched a second challenge for detecting mitosis: the MITOS–ATYPIA14 challenge. More data were made available and the tasks included also the differentiation between high and low grade nuclear atypia. Another major improvement from the first edition of the contest was that the annotations were provided by two senior pathologists and three junior pathologists. The winner of the nuclear atypia subtask was the team from University of Warwick, while the subtask of mitosis detection had two winners: the team from the Chinese university of Hong Kong for the Aperio scanner and the team from Institut Curie for the slides processed with Hamamatsu scanner. The Warwick team did not use an approach based on deep learning, but they used a generalized region covariance descriptor. The descriptor creates a geodesic measure of the manifold that represents the pixel–level feature space [70].

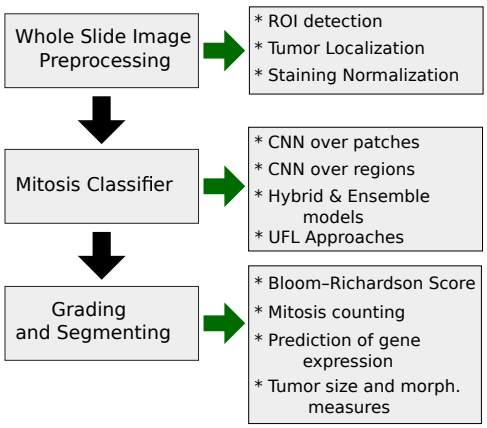


Figure 1.8 Common modules involved when applying DL to the mitosis detection problem, some of the elements overlap with those in figure 1.4. The main difference relies on the region of interest localization for patch extraction and the output of the DL classifiers.

The approach described by Chen et al. [20] is currently the method with the best performance on the 2014 MITOS-ATYPIA dataset in the task of mitosis detection. This approach is based on two steps: 1) Collecting mitosis candidates by means of a fully convolutional neural network; 2) using transfer learning to classify the candidates. The parameters of the off–the–shelf CaffeNet model were used to train three architectures with a different number of neurons in three fully connected layers. The probability of a patch being mitosis or not was computed via the average of the model decisions.

In the 2016 *MICCAI Grand Challenge on Tumor Proliferation Assessment*- TUPAC16 there were two subtasks involving mitosis detection. The first consisted of predicting a proliferation score based on mitosis counting. The second consisted of the mitosis detection of tumor regions. The results of this challenge are available on its website[10]. The detailed descriptions of the tested methods have not been uploaded yet.

Besides the challenges in mitosis detection, many researchers have been tackling this problem using deep learning techniques [20, 112, 116, 46, 111, 124]. The types of strategies can be differentiated into those involving ROI detection techniques and those that do not. Staining normalization between different microscopes/laboratories is a common pre–processing module. Further steps include tumor localization, data augmentation and boosting classifiers with difficult cases. An approach of Xu et al. is based on stacked autoencoders that create high level feature representations for 34×34 pixel patches [124]. They evaluate the approach on 37 H&E stained breast

[10] `http://tupac.tue-image.nl/node/2` as of 1 January 2017

histopathology images that were collected from a cohort of 17 patients. The result of the evaluation shows superior performance when compared with texture and morphological features. An hybrid approach is described by Wang et al. [116]. The authors propose a probabilistic fusion of two classifiers, one with handcrafted features as input and another with CNN derived features. For the difficult cases, they fuse both features to train and test a third classifier. The authors test their approach on the MITOS2012 and AMIDA13 challenge datasets showing that their method can obtain better performance than the other deep learning methods presented at the challenges.

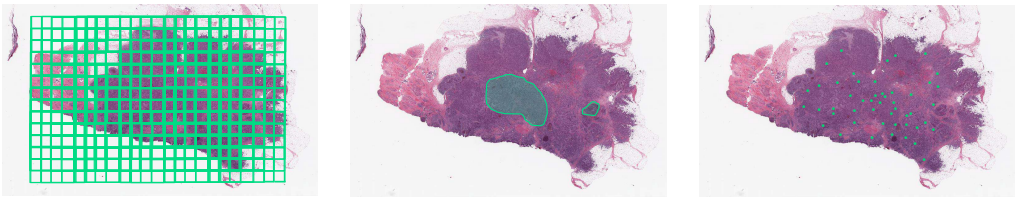
Giusti et al. [46] make a man machine comparison. The authors measure the performance of human pathologists and algorithms presented in MITOS12 [96]. They show that the most accurate pathologists were outperformed by deep learning algorithms, specifically the CNN technique from IDSIA. More recently, Veta et al. [111] go one step further in this direction and illustrate that the best measured agreement between the pathologists and an automatic DL method is comparable to the worst measured agreement among the pathologists. The authors also explain that the gap between the experts and the DL algorithm should be reduced if the researchers take into account the context of the regions in WSI and not only the patches centered at a given mitotic location.

6. Frame and Whole Slide Image Classification

To confirm or discard a diagnostic hypothesis, pathologists visually inspect whole slide images at multiple resolutions, selecting regions of interest (ROIs) and identifying structures. Image classification and disease grading in clinical routine are based on the architectural patterns in the histopathology tissue samples. In higher cancer grading, cells are poorly differentiated and prognosis is adverse with a higher risk of relapse. Pathologist grading can vary according to their personal experience and inherent subjectivity [106]. Novel approaches using deep learning are now able to classify a WSI or a region even without fully comprehending the relations between the underlying histological structures [26, 63, 65].

Different pipelines can be used to perform image classification in digital pathology. Some approaches first detect ROIs, then quantify histological primitives and finally decide the image classification. Other methods skip the middle steps entirely and decide the classification without targeting specific histological primitives. Depending on the pipeline selected, one or multiple classifiers can be trained at different resolutions and for separate histological structures. The sampling performed on the WSI can also vary according to the objectives of the proposed algorithms. Figure 1.9 presents different sampling techniques used for WSI classification.

Image classification tasks were included in medical image analysis challenges only in recent years. In *MITOS–ATYPIA–14*, high power fields from breast cancer tissue



a) Grid search b) ROIs c) Random sampling

Figure 1.9 Sampling techniques performed in WSI classification.

slides were analyzed to generate a score for nuclear atypia (i.e. nuclei shape variations compared to normal nuclei). The slides were divided into three grades according to six criteria related to nuclear pleomorphism (such as size of the nuclei, size of the nucleoli and thickness of nuclear membrane). The winning approach for this task was the one from Khan et al., both on the Aperio and Hamamatsu images [70]. This approach includes an image–level descriptor (implemented as the geodesic mean of region covariance descriptors) that enables a tractable geodesic–distance–based kNN classification using efficient kernels. This approach outperformed standard region covariance descriptors and textural descriptors for nuclear atypia.

CPM was a challenge on the classification of whole slide tissue images using two cohorts, one from patients with non small cell lung cancer and one for head and neck squamous cell carcinoma. Participants were asked to classify patients from each cohort according to their molecular subtypes. No additional information regarding the challenge results is currently available on their website[11].

CAMELYON16 was the first challenge aimed at classifying whole slide images: 270 for training and 130 for testing. The participant algorithms were implemented to discriminate between normal lymph node WSI and slides containing metastases. The probability of metastases present per image was also requested. The method by Wang et al. [115] obtained the highest score for both the WSI classification and tumor localization task. Their method consisted of a patch–based classification stage and a heatmap–based postprocessing. Positive and negative tumor samples were used for retraining a supervised classification model with the GoogLeNet architecture [104]. The predictions were embedded in a heatmap image from which 28 geometrical and morphological features were extracted to build a random forest classifier aimed at discriminating WSIs with metastases and normal WSIs.

Following the current trend in digital pathology, TUPAC16 distributed an even larger number of WSI to participants: 500 for training and 321 for testing. Unlike

[11] `http://miccai.cloudapp.net/competitions/46`, as of 1 January 2017

the binary classification from CAMELYON16, TUPAC16's main goals were to grade breast cancer WSIs into 3 classes (according to the mitosis counting among other components, i.e. Bloom & Richardson grading system) and to predict a proliferation score based on molecular data. The method from Lunit Inc. Korea was the winner for both grading tasks among the automatic methods. For the moment, there is no other information available about the methods on the challenge website [12].

7. Structure Segmentation

Defining and measuring the boundaries of tissue primitives such as glands helps in disentangling the visual variability in the WSI. This is useful to identify malignant tumors arising from glandular epithelium, which is a routinely performed task in histopathology. Structure segmentation in WSIs is a challenging task that is important for grading cancers like prostate and colon. The morphology of the glands has routinely been used by pathologists to assess the degree of malignancy of several adenocarcinomas, including prostate, breast, lung, and colon. The structure segmentation problem is now commonly addressed with many deep learning techniques such as fully convolutional neural networks (FCNN) [78], which helps to recover spatial information lost during downsampling of the usual classification architectures. Custom CNN architectures [27] have also been tested with promising results compared to the classical patch–based techniques [68].

In the 2015 MICCAI Gland Segmenting challenge (GlaS@MICCAI'2015, see Section 5), various deep learning approaches were promising regarding this direction [101]. A remarkable approach is the one of the winning group: CUMedVision[13] [21]. A modification of a fully convolutional neural network is used to segment colon tissue with a DICE score of 0.897 and an F1–score of 0.912. The approach uses *multi–level contextual features* that group features from several layers of the deep network, taking into account the intensities and varying sizes of receptive fields using a two–path FCNN.

An interesting approach for semantic segmentation in WSIs is proposed by Wang et al. [117]. They identify and tackle two main problems: the lack of training data and how to deal with the multiple sizes and shape patterns. They found that transfer learning is a reliable enhancer, both in training speed and segmentation performance. For the second problem, several FCNs are fused, varying the input size in order to be able to capture the different scales and shapes of the patterns. The approach outperforms FCN with a considerable margin in an inflammatory bowel disease dataset of H&E WSIs.

[12] as of 1 January 2017

[13] Department of Computer Science and Engineering, The Chinese University of Hong Kong

Patch–based techniques are a commonly used strategy in which the idea is that the structure to be segmented is fed in form of patches as input to the networks, for subsequently densely running of the model over the WSI. This creates a probability map for all pixels that can subsequently be used in the task of segmentation by setting a threshold in the map.

In [41] various patch–based deep learning settings are tested for both classification and segmentation of prostate cancer tissue showing promising results. Similarly, in [51], a pixel–wise classification architecture is proposed to segment prostate tissue reaching an error rate as low as 7.3% in patch classification. The problem of a better attempt in delineating the fine details in class boundaries is discussed, concluding that it is necessary to transfer these techniques to medical practice. This is an explicit trend seen in many recent methods for H&E segmentation.

In the paper of Sirinukunwattana et al. [102] a spatially constrained CNN is investigated to perform nucleus detection in colon cancer WSI. An F1 score of 0.78 is obtained classifying epithelial, inflammatory, fibroblast, and miscellaneous patches. After the model is trained, a comprehensive segmentation of a WSI is obtained. This is another example where the step of designing a particular segmentation architecture is skipped.

In the work of Drozdzal [38], the importance of the skip connections in very deep architectures for segmentation is investigated. Even though it is tested on electron microscopy images, the findings are also applicable to deep architectures that segment histopathology images. In Akram et al. [1], a two stage CNN–based method is proposed to accurately segment cells into three modalities: fluorescent imaging, phase contrast imaging and H&E images. At the first stage, a CNN is build for detecting bounding boxes. In the second stage, a CNN predicts a segmentation mask. An interesting weight sharing scheme is used that puts the ROI information into the second CNN. Xu et al. [125] present a similar approach using three CNN architectures. One is used to remove the background, another for detecting the glands and a final network to refine the edges of the mask. An object Dice score of 0.908 is achieved with respect to a FCN baseline that achieves 0.742 on the same data set.

Wenqi et al. [76] propose a hybrid approach to segment colon histology images using features from a CNN architecture and classical features like SIFT and multi–resolution local patterns among others. The performance improvement with respect to the CNN architecture alone was from 0.82 to 0.87 in Dice score. Xie et al. [123] proposed two fully convolutional regression networks to count cells in H&E images. A major component in the approach is that the training of the model is made over synthetically generated data. Nevertheless, the performance is comparable when trained on real data. Sadanandan et al. [98] also make use of data augmentation but in this case with specialized feature maps like the eigenvalue of Hessian and wavelet filtered images. Ben Taieb et al. [12] show how a CNN architecture can do joint segmen-

tation and classification over colon glands, improving the usage of both approaches separately.

The approach of Kainz et al. [68] for the GlaS15 challenge consisted of two CNN architectures regularized by a total variation criteria. The first separated benign background, benign glands, malignant background and malignant glands. A second CNN concentrated on separating very close gland objects. Tissue classification accuracies of 98% and 94% are obtained on two test sets. Ben Taieb et al. [11] encode the relationship between the different structures on the WSI with a topology aware FCNN that allows to encode geometric and topological priors of containment and detachment. This approach gained up to 10% in object–level Dice segmentation over the original FCNN approach.

Among all the work reviewed, the use of FCNN–based architectures is ubiquitous, showing good results in segmenting primitives like cells, glands and epithelium. A current trend is to include additional channels for a finer segmentation and to encode spatial, geometric and scale information in the loss functions of the models.

8. Discussion

Computer–assisted decision support of histopathology images is a challenging field that will meet the requirements of clinical practice in the near future . The fusion of traditional diagnosis with computational data analysis represents an opportunity to improve image interpretation by reducing the workload of pathologists for the most trivial tasks while also harmonizing agreement between pathologists. Since the advent of digital scanners, private companies and scientific researchers have been developing tools using machine learning that assist pathologists [52, 97]. However, assisting pathologists is not trivial due to the many technical and clinical factors that surround histopathology images. Traditional machine learning approaches obtained good results for the detection, segmentation and classification of tissue structures in the last 10–20 years. However, several technical challenges need to be solved before computer aided decision support can be applied to clinical practice. Deep learning obtained best results in most tasks in recent years and it has interesting perspectives for the future.

A first challenge in the field of histopathology CAD is represented by the high variability of tasks due to clinical characteristics of the images. Current CAD methods can obtain good results on highly specific tasks. The best performance is obtained when task–specific tuning is involved [63]. Thus, current methods are strongly sensible to the fine tuning made for a specific, limited problem. To our knowldege, there are currently no models that are capable to analyze WSI of several pathologies together.

A second challenge in the field is the development of methods that are capable of analyzing the many diverse parameters that are considered by a pathologist simultaneously. WSIs are now being used and distributed in challenges to develop algorithms

that mimic the full workflow of a pathologist. However, the workflow of pathologists often follows almost instinctive paths. While looking at an image, a pathologist zooms to different areas, searching for properties that change depending on the image and on the clinical parameters of the subject. This diagnostic workflow is partially represented by the histopathology literature, but it has not been imitated in computer science literature. The workflow involves specific challenging tasks, such as the detection of mitosis or basal cells, the recognition of glands and the classification or regions. The development of methods that are capable to perform well in each of these specific tasks are important. Combining such methods can then lead to clinical applicability.

A third challenge is represented by analysis robustness. Recent results on specific tasks can be comparable to human performances. For instance, in the challenging task of mitosis detection, the performance of deep learning based methods is now approaching the inter–observer agreement of expert pathologists [111, 22, 46]. Similarly, recent image analysis approaches can classify and grade WSI according to the tissue patterns resulting from different cancer types [26, 63] with performances that are comparable with the interpretation of pathologists on the same images. However the results are usually obtained on data sets that were usually acquired using up to 2–3 scanners, while meta–analysis comparing the same methods on data sets characterized by high heterogeneity have not been performed yet. Thus, the challenge for the performance of each algorithm is to make it robust on different whole slide images.

Structure segmentation in WSI is a challenging task that is important for grading some cancers. The morphology of glands is routinely used by pathologists to assess the degree of malignancy of several adenocarcinomas, including prostate, breast, lung, and colon cancer. FCNN–based architectures are frequent and promising for cells, glands and epithelium segmentation, while current trends include the use of additional channels to benefit of spatial, geometric and scale information. However, also for this task, robust and generalizable results must be obtained before translating them into practice.

Future perspectives and trends for histopathology CAD systems include the capability to better interact with images from different scanners and across pathologies, as well as the creation of methods that can learn from unlabelled or weakly labelled data. During pre–processing, advanced normalization methods can increase image similarity, thus making the CAD methods more reliable even on images acquired with different scanners and with varying staining procedures. The creation of extended data sets that include numerous pathologies may allow the creation of algorithms that rely on a deep comprehension of the whole histopathology field, thus making CAD systems capable of working in several anatomical structures reliably. Finally, the recent advent and success of deep learning methods in digital pathology suggests that this quickly improve the performance of CAD systems and accelerate their use, possible via the development of unsupervised learning methods. Most of current learning algorithms

require manually annotated training sets. This is particularly evident for deep learning methods that usually require large amounts of data to work well. It is often difficult to obtain large amount of manually annotated WSIs because the task is tedious, it requires a medical background and pathologists are usually not very interested in doing it. Unsupervised learning methods (as in [49]) can help to skip the data annotation phase, thus allowing to exploit histopathology images automatically.

In conclusion, currently WSI and deep learning are revolutionizing histopathology CAD, and they may soon help to decrease the work load of pathologists for the most trivial tasks. This should allow pathologists to concentrate on the most difficult cases and result in an even deeper understanding of the pathologic processes via the machine learning approaches.

Acknowledgments

his work was supported by the Eurostars project 9653 SLDESUTO–BOX.

References

1. Akram, S. U., Kannala, J., Eklund, L., and Heikkilä, J. Cell segmentation proposal network for microscopy image analysis. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis* (2016), Springer, pp. 21–29.
2. Al-Kofahi, Y., Lassoued, W., Lee, W., and Roysam, B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering 57*, 4 (2010), 841–852.
3. Alexandratou, E., Atlamazoglou, V., Thireou, T., Agrogiannis, G., Togas, D., Kavantzas, N., Patsouris, E., and Yova, D. Evaluation of machine learning techniques for prostate cancer diagnosis and gleason grading. *International Journal of Computational Intelligence in Bioinformatics and Systems Biology 1*, 3 (2010), 297–315.
4. Altunbay, D., Cigir, C., Sokmensuer, C., and Gunduz-Demir, C. Color graphs for automated cancer diagnosis and grading. *IEEE Transactions on Biomedical Engineering 57*, 3 (2010), 665–674.
5. Aptoula, E., Courty, N., and Lefèvre, S. Mitosis detection in breast cancer histological images with mathematical morphology. In *Signal Processing and Communications Applications Conference (SIU), 2013 21st* (2013), IEEE, pp. 1–4.
6. Arevalo, J., Cruz-Roa, A., Arias, V., Romero, E., and González, F. A. An unsupervised feature learning framework for basal cell carcinoma image analysis. *Artificial intelligence in medicine 64*, 2 (2015), 131–145.
7. Arevalo, J., Cruz-Roa, A., et al. Histopathology image representation for automatic analysis: a state-of-the-art review. *Revista Med 22*, 2 (2014), 79–91.
8. Arevalo, J., Cruz-Roa, A., and González, F. A. Hybrid image representation learning model with invariant features for basal cell carcinoma detection. In *IX international seminar on medical information processing and analysis* (2013), International Society for Optics and Photonics, pp. 89220M–89220M.
9. Bartels, P., Thompson, D., Bibbo, M., and Weber, J. Bayesian belief networks in quantitative histopathology. *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology 14*, 6 (1992), 459–473.
10. Bengio, Y., Courville, A. C., and Vincent, P. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR, abs/1206.5538 1* (2012).
11. BenTaieb, A., and Hamarneh, G. Topology aware fully convolutional networks for histology gland segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2016), Springer, pp. 460–468.
12. BenTaieb, A., Kawahara, J., and Hamarneh, G. Multi-loss convolutional networks for gland analysis in microscopy. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on* (2016), IEEE, pp. 642–645.
13. Bloom, H., and Richardson, W. Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. *British journal of cancer 11*, 3 (1957), 359.
14. Boucheron, L. E. *Object-and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer*. University of California at Santa Barbara, 2008.
15. Burt, P., and Adelson, E. The laplacian pyramid as a compact image code. *IEEE Transactions on communications 31*, 4 (1983), 532–540.
16. Can, A., Bello, M., Cline, H. E., Tao, X., Ginty, F., Sood, A., Gerdes, M., and Montalto, M. Multi-modal imaging of histological tissue sections. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2008), IEEE, pp. 288–291.
17. Chang, H., Zhou, Y., Borowsky, A., Barner, K., Spellman, P., and Parvin, B. Stacked predictive sparse decomposition for classification of histology sections. *International Journal of Computer Vision 113*, 1 (2015), 3–18.
18. Chekkoury, A., Khurd, P., Ni, J., Bahlmann, C., Kamen, A., Patel, A., Grady, L., Singh, M., Groher, M., Navab, N., et al. Automated malignancy detection in breast histopathological images. In *SPIE Medical Imaging* (2012), International Society for Optics and Photonics, pp. 831515–831515.
19. Chen, C. L., Mahjoubfar, A., Tai, L.-C., Blaby, I. K., Huang, A., Niazi, K. R., and Jalali, B. Deep

learning in label-free cell classification. *Scientific reports 6* (2016).

20. Chen, H., Dou, Q., Wang, X., Qin, J., and Heng, P. A. Mitosis detection in breast cancer histology images via deep cascaded networks. In *Thirtieth AAAI Conference on Artificial Intelligence* (2016).
21. Chen, H., Qi, X., Yu, L., and Heng, P.-A. Dcan: Deep contour-aware networks for accurate gland segmentation. *arXiv preprint arXiv:1604.02677* (2016).
22. Chen, T., and Chefd'hotel, C. Deep learning based automatic immune cell detection for immunohistochemistry images. In *International Workshop on Machine Learning in Medical Imaging* (2014), Springer, pp. 17–24.
23. Cireşan, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J. Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer-assisted Intervention* (2013), Springer, pp. 411–418.
24. Cosatto, E., Miller, M., Graf, H. P., and Meyer, J. S. Grading nuclear pleomorphism on histological micrographs. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (2008), IEEE, pp. 1–4.
25. Cruz-Roa, A., Arevalo, J., Basavanhally, A., Madabhushi, A., and González, F. A comparative evaluation of supervised and unsupervised representation learning approaches for anaplastic medulloblastoma differentiation. In *Tenth International Symposium on Medical Information Processing and Analysis* (2015), International Society for Optics and Photonics, pp. 92870G–92870G.
26. Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., and Madabhushi, A. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *SPIE medical imaging* (2014), International Society for Optics and Photonics, pp. 904103–904103.
27. Cruz-Roa, A. A., Ovalle, J. E. A., Madabhushi, A., and Osorio, F. A. G. 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* (2013), Springer, pp. 403–410.
28. Dalle, J.-R., Li, H., Huang, C.-H., Leow, W. K., Racoceanu, D., and Putti, T. C. Nuclear pleomorphism scoring by selective cell nuclei detection. In *WACV* (2009).
29. Demir, C., and Yener, B. Automated cancer diagnosis based on histopathological images: a systematic survey. *Rensselaer Polytechnic Institute, Tech. Rep* (2005).
30. Deng, L., Hinton, G., and Kingsbury, B. New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), IEEE, pp. 8599–8603.
31. Depeursinge, A., Foncubierta-Rodríguez, A., Van De Ville, D., and Müller, H. Three-dimensional solid texture analysis and retrieval in biomedical imaging: review and opportunities. *Medical Image Analysis 18*, 1 (2014), 176–196.
32. Depeursinge, A., and Müller, H. Fusion techniques for combining textual and visual information retrieval. In *ImageCLEF*, H. Müller, P. Clough, T. Deselaers, and B. Caputo, Eds., vol. 32 of *The Springer International Series On Information Retrieval*. Springer Berlin Heidelberg, 2010, pp. 95–114.
33. Diamond, J., Anderson, N. H., Bartels, P. H., Montironi, R., and Hamilton, P. W. The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia. *Human Pathology 35*, 9 (2004), 1121–1131.
34. DiFranco, M. D., O'Hurley, G., Kay, E. W., Watson, R. W. G., and Cunningham, P. Ensemble based system for whole-slide prostate cancer probability mapping using color texture features. *Computerized medical imaging and graphics 35*, 7 (2011), 629–645.
35. Doyle, S., Hwang, M., Shah, K., Madabhushi, A., Feldman, M., and Tomaszeweski, J. Automated grading of prostate cancer using architectural and textural image features. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2007), IEEE, pp. 1284–1287.
36. Doyle, S., Madabhushi, A., Feldman, M., and Tomaszeweski, J. A boosting cascade for automated detection of prostate cancer from digitized histology. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2006), Springer, pp. 504–511.
37. Doyle, S., Rodriguez, C., Madabhushi, A., Tomaszeweski, J., and Feldman, M. Detecting prostatic

adenocarcinoma from digitized histology using a multi-scale hierarchical classification approach. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE* (2006), IEEE, pp. 4759–4762.

38. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., and Pal, C. The importance of skip connections in biomedical image segmentation. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis* (2016), Springer, pp. 179–187.
39. Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., and Bengio, S. Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research 11*, Feb (2010), 625–660.
40. Fine, J. L., Grzybicki, D. M., Silowash, R., Ho, J., Gilbertson, J. R., Anthony, L., Wilson, R., Parwani, A. V., Bastacky, S. I., Epstein, J. I., et al. Evaluation of whole slide image immunohistochemistry interpretation in challenging prostate needle biopsies. *Human pathology 39*, 4 (2008), 564–572.
41. Flood, G. Deep learning with a dag structure for segmentation and classification of prostate cancer. *Master's Theses in Mathematical Sciences* (2016).
42. Freund, Y., and Schapire, R. E. A desicion-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory* (1995), Springer, pp. 23–37.
43. Fu, H., Qiu, G., Shu, J., and Ilyas, M. A novel polar space random field model for the detection of glandular structures. *IEEE transactions on medical imaging 33*, 3 (2014), 764–776.
44. Ghaznavi, F., Evans, A., Madabhushi, A., and Feldman, M. Digital imaging in pathology: whole-slide imaging and beyond. *Annual Review of Pathology: Mechanisms of Disease 8* (2013), 331–359.
45. Ghosh, J. Multiclassifier systems: Back to the future. In *International Workshop on Multiple Classifier Systems* (2002), Springer, pp. 1–15.
46. Giusti, A., Caccia, C., Cireşari, D. C., Schmidhuber, J., and Gambardella, L. M. A comparison of algorithms and humans for mitosis detection. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (2014), IEEE, pp. 1360–1363.
47. Gleason, D. F. Histologic grading of prostate cancer: a perspective. *Human pathology 23*, 3 (1992), 273–279.
48. Glorot, X., Bordes, A., and Bengio, Y. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (2011), pp. 513–520.
49. Goodfellow, I., Bengio, Y., and Courville, A. Deep learning. Book in preparation for MIT Press, 2016.
50. Greenspan, H., van Ginneken, B., and Summers, R. M. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging 35*, 5 (2016), 1153–1159.
51. Gummeson, A. Prostate cancer classification using convolutional neural networks. *Master's Theses in Mathematical Sciences* (2016).
52. Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., and Yener, B. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering 2* (2009), 147–171.
53. Gutman, D., Codella, N. C. F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., and Halpern, A. 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* (2016).
54. Han, J., Fontenay, G. V., Wang, Y., Mao, J.-H., and Chang, H. Phenotypic characterization of breast invasive carcinoma via transferable tissue morphometric patterns learned from glioblastoma multiforme. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on* (2016), IEEE, pp. 1025–1028.
55. Ho, J., Parwani, A. V., Jukic, D. M., Yagi, Y., Anthony, L., and Gilbertson, J. R. Use of whole slide imaging in surgical pathology quality assurance: design and pilot validation studies. *Human pathology 37*, 3 (2006), 322–331.

56. Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E., and Saltz, J. H. Efficient multiple instance convolutional neural networks for gigapixel resolution image classification. *arXiv preprint* (2015).
57. Huang, P.-W., and Lai, Y.-H. Effective segmentation and classification for hcc biopsy images. *Pattern Recognition 43*, 4 (2010), 1550–1563.
58. Huang, P.-W., and Lee, C.-H. Automatic classification for pathological prostate images based on fractal analysis. *IEEE transactions on medical imaging 28*, 7 (2009), 1037–1050.
59. Huisman, A., Looijen, A., van den Brink, S. M., and van Diest, P. J. Creation of a fully digital pathology slide archive by high-volume tissue slide scanning. *Human pathology 41*, 5 (2010), 751–757.
60. Irshad, H., et al. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *Journal of pathology informatics 4*, 1 (2013), 10.
61. Irshad, H., Veillard, A., Roux, L., and Racoceanu, D. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review–current status and future potential. *IEEE reviews in biomedical engineering 7* (2014), 97–114.
62. Jafari-Khouzani, K., and Soltanian-Zadeh, H. Multiwavelet grading of pathological images of prostate. *IEEE Transactions on Biomedical Engineering 50*, 6 (2003), 697–704.
63. Janowczyk, A., and Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics 7* (2016).
64. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (2014), ACM, pp. 675–678.
65. Jimenez-del Toro, O., Atzori, M., Andersson, M., Eurén, K., Hedlund, M., Rönnquist, P., and Müller, H. Convolutional neural networks for an automatic classification of prostate tissue slides with high—grade gleason score. In *SPIE Medical Imaging* (2017), International Society for Optics and Photonics.
66. Jolliffe, I. *Principal component analysis*. Wiley Online Library, 2002.
67. Jung, C., and Kim, C. Segmenting clustered nuclei using h-minima transform-based marker extraction and contour parameterization. *IEEE transactions on biomedical engineering 57*, 10 (2010), 2600–2604.
68. Kainz, P., Pfeiffer, M., and Urschler, M. Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation. *arXiv preprint arXiv:1511.06919* (2015).
69. Khan, A. M., Rajpoot, N., Treanor, D., and Magee, D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Transactions on Biomedical Engineering 61*, 6 (2014), 1729–1738.
70. Khan, A. M., Sirinukunwattana, K., and Rajpoot, N. A global covariance descriptor for nuclear atypia scoring in breast histopathology images. *IEEE Journal of Biomedical and Health Informatics 19*, 5 (2015), 1637–1647.
71. Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H., and Gurcan, M. N. Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation. *Pattern Recognition 42*, 6 (2009), 1080–1092.
72. Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.
73. LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. *Nature 521*, 7553 (2015), 436–444.
74. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278–2324.
75. Leong, F. W., Brady, M., and McGee, J. O. Correction of uneven illumination (vignetting) in digital microscopy images. *Journal of clinical pathology 56*, 8 (2003), 619–621.
76. Li, W., Manivannan, S., Akbar, S., Zhang, J., Trucco, E., and McKenna, S. J. Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on* (2016), IEEE, pp. 1405–1408.
77. Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-van de Kaa, C., Bult, P., van Ginneken, B., and van der Laak, J. Deep learning as a tool for

increased accuracy and efficiency of histopathological diagnosis. *Scientific reports 6* (2016).

78. Long, J., Shelhamer, E., and Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 3431–3440.
79. Lyon, H. O., De Leenheer, A., Horobin, R., Lambert, W., Schulte, E., Van Liedekerke, B., and Wittekind, D. Standardization of reagents and methods used in cytological and histological practice with emphasis on dyes, stains and chromogenic reagents. *The Histochemical Journal 26*, 7 (1994), 533–544.
80. Magee, D., Treanor, D., Crellin, D., Shires, M., Smith, K., Mohee, K., and Quirke, P. Colour normalisation in digital histopathology images. In *Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)* (2009), vol. 100, Citeseer.
81. Malon, C. D., Cosatto, E., et al. Classification of mitotic figures with convolutional neural networks and seeded blob features. *Journal of pathology informatics 4*, 1 (2013), 9.
82. Manion, E., Cohen, M. B., and Weydert, J. Mandatory second opinion in surgical pathology referral material: clinical consequences of major disagreements. *The American journal of surgical pathology 32*, 5 (2008), 732–737.
83. Marty, G. D., et al. Blank-field correction for achieving a uniform white background in brightfield digital photomicrographs. *BioTechniques 42*, 6 (2007), 716.
84. Naik, S., Doyle, S., Feldman, M., Tomaszewski, J., and Madabhushi, A. Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information. In *MIAAB workshop* (2007), Citeseer, pp. 1–8.
85. Nayak, N., Chang, H., Borowsky, A., Spellman, P., and Parvin, B. Classification of tumor histopathology via sparse feature learning. In *2013 IEEE 10th International Symposium on Biomedical Imaging* (2013), IEEE, pp. 410–413.
86. Noël, H., Roux, L., Lu, S., and Boudier, T. Detection of high-grade atypia nuclei in breast cancer imaging. In *SPIE Medical Imaging* (2015), International Society for Optics and Photonics, pp. 94200R–94200R.
87. Otálora, S., Cruz-Roa, A., Arevalo, J., Atzori, M., Madabhushi, A., Judkins, A. R., González, F., Müller, H., and Depeursinge, A. Combining unsupervised feature learning and riesz wavelets for histopathology image representation: Application to identifying anaplastic medulloblastoma. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2015), Springer, pp. 581–588.
88. Pawlowski, N., Caicedo, J. C., Singh, S., Carpenter, A. E., and Storkey, A. Automating morphological profiling with generic deep convolutional networks. *bioRxiv* (2016).
89. Perkins, S., Lacker, K., and Theiler, J. Grafting: Fast, incremental feature selection by gradient descent in function space. *Journal of machine learning research 3*, Mar (2003), 1333–1356.
90. Petushi, S., Garcia, F. U., Haber, M. M., Katsinis, C., and Tozeren, A. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC medical imaging 6*, 1 (2006), 1.
91. Piccinini, F., Lucarelli, E., Gherardi, A., and Bevilacqua, A. Multi-image based method to correct vignetting effect in light microscopy images. *Journal of microscopy 248*, 1 (2012), 6–22.
92. Pudil, P., Novovičová, J., and Kittler, J. Floating search methods in feature selection. *Pattern recognition letters 15*, 11 (1994), 1119–1125.
93. Qureshi, H., Sertel, O., Rajpoot, N., Wilson, R., and Gurcan, M. Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2008), Springer, pp. 196–204.
94. Rajpoot, K., and Rajpoot, N. Svm optimization for hyperspectral colon tissue cell classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2004), Springer, pp. 829–837.
95. Romo-Bucheli David, Janowczyk Andrew, Gilmore Hannah, Romero Eduardo, and Madabhushi Anant. Automated Tubule Nuclei Quantification and Correlation with Oncotype DX risk categories in ER+ Breast Cancer Whole Slide Images. *Scientific Reports 6* (sep 2016), 32706.

96. Roux, L., Racoceanu, D., Loménie, N., Kulikova, M., Irshad, H., Klossa, J., Capron, F., Genestie, C., Naour, G., and Gurcan, M. Mitosis detection in breast cancer histological images An ICPR 2012 contest. *Journal of Pathology Informatics 4*, 1 (2013), 8.
97. Rubin, R., Strayer, D. S., Rubin, E., et al. *Rubin's pathology: clinicopathologic foundations of medicine*. Lippincott Williams & Wilkins, 2008.
98. Sadanandan, S. K., Ranefall, P., and Wählby, C. Feature augmented deep neural networks for segmentation of cells. In *European Conference on Computer Vision* (2016), Springer, pp. 231–243.
99. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks 61* (2015), 85–117.
100. Shah, M., Rubadue, C., Suster, D., and Wang, D. Deep learning assessment of tumor proliferation in breast cancer histological images. *arXiv preprint arXiv:1610.03467* (2016).
101. Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P.-A., Guo, Y. B., Wang, L. Y., Matuszewski, B. J., Bruni, E., Sanchez, U., et al. Gland segmentation in colon histology images: The glas challenge contest. *arXiv preprint arXiv:1603.00275* (2016).
102. Sirinukunwattana, K., Raza, S. E. A., Tsang, Y.-W., Snead, D. R., Cree, I. A., and Rajpoot, N. M. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging 35*, 5 (2016), 1196–1206.
103. Sirinukunwattana, K., Snead, D. R., and Rajpoot, N. M. A stochastic polygons model for glandular structures in colon histology images. *IEEE transactions on medical imaging 34*, 11 (2015), 2366–2378.
104. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 1–9.
105. Ta, V.-T., Lézoray, O., Elmoataz, A., and Schüpp, S. Graph-based tools for microscopic cellular image segmentation. *Pattern Recognition 42*, 6 (2009), 1113–1125.
106. Trpkov, K. Contemporary gleason grading system. In *Genitourinary Pathology*. Springer, 2015, pp. 13–32.
107. Vanegas, J. A., Arevalo, J., and González, F. A. Unsupervised feature learning for content-based histopathology image retrieval. In *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)* (2014), IEEE, pp. 1–6.
108. Velez, N., Jukic, D., and Ho, J. Evaluation of 2 whole-slide imaging applications in dermatopathology. *Human pathology 39*, 9 (2008), 1341–1349.
109. Veta, M., Huisman, A., Viergever, M. A., van Diest, P. J., and Pluim, J. P. Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2011), IEEE, pp. 618–621.
110. Veta, M., Pluim, J. P., van Diest, P. J., and Viergever, M. A. Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering 61*, 5 (2014), 1400–1411.
111. Veta, M., van Diest, P. J., Jiwa, M., Al-Janabi, S., and Pluim, J. P. Mitosis counting in breast cancer: Object-level interobserver agreement and comparison to an automatic method. *PloS one 11*, 8 (2016), e0161286.
112. Veta, M., van Diest, P. J., Willems, S. M., Wang, H., Madabhushi, A., Cruz-Roa, A., Gonzalez, F., Larsen, A. B. L., Vestergaard, J. S., Dahl, A. B., Cireşan, D. C., Schmidhuber, J., Giusti, A., Gambardella, L. M., Tek, F. B., Walter, T., Wang, C.-W., Kondo, S., Matuszewski, B. J., Precioso, F., Snell, V., Kittler, J., de Campos, T. E., Khan, A. M., Rajpoot, N. M., Arkoumani, E., Lacle, M. M., Viergever, M. A., and Pluim, J. P. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical image analysis 20*, 1 (2015), 237–248.
113. Vink, J., Van Leeuwen, M., Van Deurzen, C., and De Haan, G. Efficient nucleus detector in histopathology images. *Journal of microscopy 249*, 2 (2013), 124–135.
114. Wählby, C., SINTORN, I.-M., Erlandsson, F., Borgefors, G., and Bengtsson, E. Combining intensity, edge and shape information for 2d and 3d segmentation of cell nuclei in tissue sections. *Journal of Microscopy 215*, 1 (2004), 67–76.
115. Wang, D., Khosla, A., Gargeya, R., Irshad, H., and Beck, A. H. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718* (2016).

116. Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F., and Madabhushi, A. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging 1*, 3 (2014), 034003–034003.
117. Wang, J., MacKenzie, J. D., Ramachandran, R., and Chen, D. Z. A deep learning approach for semantic segmentation in histology tissue images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2016), Springer, pp. 176–184.
118. Washington, M. K., Berlin, J., Branton, P., Burgart, L. J., Carter, D. K., Fitzgibbons, P. L., Halling, K., Frankel, W., Jessup, J., Kakar, S., et al. Protocol for the examination of specimens from patients with primary carcinoma of the colon and rectum. *Archives of pathology & laboratory medicine 133*, 10 (2009), 1539–1551.
119. Weind, K. L., Maier, C. F., Rutt, B. K., and Moussa, M. Invasive carcinomas and fibroadenomas of the breast: comparison of microvessel distributions-implications for imaging modalities. *Radiology 208*, 2 (1998), 477–483.
120. Weinstein, R. S., Graham, A. R., Richter, L. C., Barker, G. P., Krupinski, E. A., Lopez, A. M., Erps, K. A., Bhattacharyya, A. K., Yagi, Y., and Gilbertson, J. R. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Human pathology 40*, 8 (2009), 1057–1069.
121. Woolgar, J. A., Ferlito, A., Devaney, K. O., Rinaldo, A., and Barnes, L. How trustworthy is a diagnosis in head and neck surgical pathology? a consideration of diagnostic discrepancies (errors). *European Archives of Oto-Rhino-Laryngology 268*, 5 (2011), 643–651.
122. WU, H.-S., Xu, R., Harpaz, N., Burstein, D., and Gil, J. Segmentation of intestinal gland images with iterative region growing. *Journal of Microscopy 220*, 3 (2005), 190–204.
123. Xie, W., Noble, J. A., and Zisserman, A. Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* (2016), 1–10.
124. Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., and Madabhushi, A. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. *IEEE transactions on medical imaging 35*, 1 (2016), 119–130.
125. Xu, Y., Li, Y., Liu, M., Wang, Y., Fan, Y., Lai, M., Chang, E. I., et al. Gland instance segmentation by deep multichannel neural networks. *arXiv preprint arXiv:1607.04889* (2016).
126. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., Eric, I., and Chang, C. 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)* (2014), IEEE, pp. 1626–1630.