Institute for Molecular Medicine Finland (FIMM)
Helsinki Institute of Life Science (HiLIFE)
Faculty of Medicine
Doctoral Programme in Biomedicine
University of Helsinki

COMPUTER VISION FOR TISSUE CHARACTERIZATION AND OUTCOME PREDICTION IN CANCER

Riku Turkki, MSc (Tech.)

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University Helsinki, for public examination in Auditorium XV, Fabianinkatu 33, University Main Building, on 24th of August 2018, at 12 noon.

Helsinki, 2018

Supervised by

Docent Johan Lundin, MD, PhD
Institute for Molecular Medicine Finland (FIMM),
University of Helsinki,
Helsinki, Finland

Docent Nina Linder, MD, PhD
Institute for Molecular Medicine Finland (FIMM),
University of Helsinki,
Helsinki, Finland

Thesis committee

Professor Jorma Isola, MD, PhD
BioMediTech,
University of Tampere,
Tampere, Finland

Docent Jorma Laaksonen, DTech
Department of Computer Science,
Aalto University School of Science,
Espoo, Finland

Reviewed by

Associate Professor Claes Lundström, PhD,
Center for Medical Image Analysis and Visualization
Linköping University,
Linköping, Sweden

Associate Professor Johan Hartman, MD, PhD
Department of Oncology-Pathology,
Karolinska Institutet,
Stockholm, Sweden

Opponent

Docent Pekka Ruusuvuori, DTech
BioMediTech
University of Tampere
Tampere, Finland

Custos

Professor Sampsa Hautaniemi, DTech
Genome-Scale Biology Research Program,
Medical Faculty, University of Helsinki,
Helsinki, Finland

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam
Universitatis Helsinkiensis (50/2018)

ISBN 978-951-51-4397-6 (Print)
ISBN 978-951-51-4398-3 (Online)
ISSN 2343-3161 (Print)
ISSN 2343-317X (Online)
Unigrafia, Helsinki 2018

To my family

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I. **Turkki R**, Linder N, Holopainen T, Wang Y, Grote A, Lundin M, Alitalo K & Lundin J "Assessment of tumour viability in human lung cancer xenografts with texture-based image analysis" *Journal of Clinical Pathology*, 68:614-621, 2015

II. **Turkki R**, Linder N, Kovanen PE, Pellinen T & Lundin J "Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples" *Journal of Pathology Informatics*, 7:38, 2016

III. **Turkki R**, Byckhov D, Lundin M, Isola J, Nordling S, Kovanen PE, Verrill C, von Smitten K, Joensuu H, Lundin J & Linder N "Breast cancer outcome prediction with tumour tissue images and machine learning" *Manuscript*, 2018

The publications are referred to in the text by their roman numerals. The original publications are reprinted with the permission of their copyright holders.

ABBREVIATIONS

κ	kappa-value
95% CI	95% confidence interval
AUCROC	area under receiver operating characteristics curve
CAD	computer-aided diagnosis
CNN	convolutional neural network
DRS	digital risk score
DSS	disease-specific survival
ECW	enhanced compressed wavelet
ER	estrogen receptor
FFPE	formalin-fixed paraffin-embedded
FIMM	Institute for Molecular Medicine Finland
FOV	field of view
FV	fisher vector
GLCM	gray-level co-occurrence matrix
GMM	gaussian mixture model
H&E	hematoxylin and eosin
HER2	human epidermal growth factor receptor 2
HR	hazard ratio
IFV	improved Fisher vector
IHC	immunohistochemistry
LBP	local binary pattern
NSCLC	non-small cell lung cancer
OS	overall survival
p	number of sampling points in a texture feature pattern
P	p-value
PCA	principal component analysis
PR	progesterone receptor
r	radius of texture feature pattern
riu2	rotation invariant 2-uniform
ROCAUC	area under receiver operating characteristics curve
STEI	single tissue entity image
SVM	support vector machine
TIL	tumor-infiltrating lymphocyte
TMA	tissue microarray
VAR	rotation invariant variance
WSI	whole-slide image

ABSTRACT

The aim of this dissertation was to investigate the use of computer vision for tissue characterization and patient outcome prediction in cancer. This work focused on analysis of digitized tissue specimens, which were stained only for basic morphology (i.e. hematoxylin and eosin). The applicability of texture analysis and convolutional neural networks was evaluated for detection of biologically and clinically relevant features. Moreover, novel approaches to guide ground-truth annotation and outcome-supervised learning for prediction of patient survival directly from the tumor tissue images without expert guidance was investigated.

We first studied quantification of tumor viability through segmentation of necrotic and viable tissue compartments. We developed a regional texture analysis method, which was trained and tested on whole sections of mouse xenograft models of human lung cancer. Our experiments showed that the proposed segmentation was able to discriminate between viable and non-viable tissue regions with high accuracy when compared to human expert assessment.

We next investigated the feasibility of pre-trained convolutional neural networks in analysis of breast cancer tissue, aiming to quantify tumor-infiltrating lymphocytes in the specimens. Interestingly, our results showed that pre-trained convolutional neural networks can be adapted for analysis of histological image data, outperforming texture analysis. The results also indicated that the computerized assessment was on par with pathologist assessments. Moreover, the study presented an image annotation technique guided by specific antibody staining for improved ground-truth labeling.

Direct outcome prediction in breast cancer was then studied using a nationwide patient cohort. A computerized pipeline, which incorporated orderless feature aggregation and convolutional image descriptors for outcome-supervised classification, resulted in a risk grouping that was predictive of both disease-specific and overall survival. Surprisingly, further analysis suggested that the computerized risk prediction was also an independent prognostic factor that provided information complementary to the standard clinicopathological factors.

This doctoral thesis demonstrated how computer-vision methods can be powerful tools in analysis of cancer tissue samples, highlighting strategies for supervised characterization of tissue entities and an approach for identification of novel prognostic morphological features.

1 INTRODUCTION

Despite improved understanding of the molecular characteristics of cancer, histological analysis of tumor specimens continues to have a key role in diagnosis and outcome prediction of cancer. For instance, a pathologist's evaluation of tumor morphology and series of tissue entities have an important role in determining what treatment options are best suited for a patient, and what is the likelihood that the disease will return. However, manual histological evaluation of cancer tissue is poorly reproducible and only semi-quantitative (Vestjens et al., 2012). Moreover, the evaluations are often time-consuming and labor-intensive.

Recent technological advances in digital pathology have allowed large-scale and high-precision digitization of tissue specimens (Pantanowitz et al., 2011). In parallel, computer vision, supplemented with machine learning, has enabled unprecedented accuracy for mining information in images (LeCun et al., 2015). Thereby, computer-vision methods are increasingly adapted to histological analysis of cancer tissue. These novel methods have the potential to enable more quantitative and reproducible analysis of tissue specimens (Djuric et al., 2017). In addition, computerized analysis of cancer tissue specimens may lower the pathologists' workload and thus decrease time needed for diagnosis.

To develop and identify computer vision methods that can be utilized in analysis of histological cancer specimens, we studied tissue characterization and patient outcome prediction. Tumor tissue is composed of various entities that hold clinically important information on the disease. We studied computerized quantification of two tumor entities, namely necrosis and tumor-infiltrating lymphocytes. Furthermore, computerized methods may be capable of discovering novel risk groups in large patient cohorts. To this end, we investigated direct outcome prediction using cancer tissue images as an input and patient survival data as the endpoint.

2 REVIEW OF THE LITERATURE

2.1 Cancer histopathology

Histopathology (or histology) of cancer is the study of tumor tissue through a light microscope (Weinberg, 2007). Histologic evaluation of tumor tissue regularly serves as the gold standard for cancer diagnosis and is one of the principal determinants in patient outcome prediction and therapeutic decision making (Chan, 2014).

2.1.1 Histological assessments

Histological assessment of a tumor tissue facilitates patient stratification into subtypes based on the specimens' morphological features and biomarker expression status. Histological grade and type are the principal measurements of morphological features, whereas immunohistochemistry (IHC) is used for assessment of specific biomarkers. (Fletcher, 2013)

Histological grade is a measurement of tumor differentiation and is assessed from hematoxylin and eosin (H&E)-stained tumor specimens. Low grade indicates that a tumor is well differentiated, or that the cells and tissue structures resemble the cells and structures in normal, non-cancerous tissue (Elston & Ellis, 1991; Epstein et al., 2016). Higher grade tumors are less differentiated and they differ more from normal tissue morphology. In general, higher grade tumors are more aggressive and are likely to metastasize. Accordingly, patients with higher grade tumors have a less favorable prognosis (Meyer et al., 2005; Sun et al., 2006). Depending on the cancer type, attributes of different tissue entities are considered in grading. For instance, in prostate cancer the grading is based on the Gleason score that evaluates tumor histologic patterns (Epstein et al., 2016). On the other hand, three well-defined tissue entities (i.e. tubular differentiation, nuclear pleomorphism, and mitotic count) are considered in breast cancer (Elston & Ellis, 1991).

Histological tumor type is likewise assessed from H&E-stained tissue and is a classification based on which tissue the cancer

originates from. The majority of cancers are classified as carcinomas, indicating that the cancer cells originated from epithelial tissue (Weinberg, 2007). Moreover, morphological features such as growth patterns and structures that cancer cells form facilitate more detailed histological subtyping. The presence of entities such as necrosis, immune cells, vessels, and amount and features of stroma contribute to the histological type. Tumors present highly heterogeneous histologies and may be mixtures of known types. For example, the WHO classification of breast tumors describes at least 17 different histological subtypes with distinctive features (Tavassoéli & Devilee, 2003).

By definition, biomarkers are measurements that indicate a state of a disease (Strimbu & Tavel, 2010). In histological analysis of cancer, this usually refers to detection of proteins or amino acids that are either predictive or prognostic using IHC (Matos et al., 2010). A predictive biomarker is a measurement that has an association with patient response to a specific treatment, whereas a prognostic factor is associated with patient outcome regardless of therapy (Oldenhuis et al., 2008). For instance, in breast cancer steroid hormone receptors are important biomarkers and are therefore assessed for complete diagnosis to support histological grade and type (Nicolini et al., 2017).

2.1.2 Tissue preparation

To prevent tissue degradation and to preserve the morphological and molecular composition, the removed tissue specimens require specific preparation. The first step in tissue preparation is chemical fixation by immersing the tissue in a formaldehyde solution (also known as formalin). The tissue specimens are next dehydrated in a series of alcohol baths and cleared with xylene, after which they are infiltrated and embedded in paraffin. Finally, the formalin-fixed, paraffin-embedded (FFPE) specimen block is cut with a microtome into thin sections (3-7 μm) that are mounted onto glass microscope slides. (Junqueira & Carneiro, 2005).

Tissue microarray (TMA) is a technique for constructing multi-specimen paraffin blocks (Kononen et al., 1998). Needle biopsies (0.6-1.0 mm) are punched from prepared FFPE blocks and transferred to a

recipient block in an array pattern. TMAs allow for simultaneous analysis of up to 1,000 individual patients.

2.1.3 Staining

A thin tissue section is nearly transparent and therefore different staining methods are used to provide contrast for the morphological structures or to highlight specific entities, such as proteins (Weinberg, 2007).

H&E has been the principal staining in histology for over a century (Chan, 2014). This staining highlights the details of tissues and cells and provides the contrast required for visual or computerized interpretation. Hematoxylin colors the basophilic tissue components (such as cell nuclei) dark blue to violet, whereas eosin provides varied shades of red, pink, and orange to the cytoplasm and extracellular proteins (Chan, 2014).

IHC is a technique based on an antigen-antibody reaction that is used to visualize and localize specific macromolecules (such as proteins and amino acids) within tissues (Coons et al., 1941). The antigen-antibody binding reaction is visualized with a chromogenic or a fluorescent staining method and detected with a microscope. IHC technologies allow for subcellular detection of target molecules and can be therefore utilized for visualizing individual cells or cell populations of interest.

2.2 Studied histological assessments

2.2.1 Tumor necrosis

Necrosis describes cell death that is usually caused by external factors such as trauma or extreme conditions (Robbins et al., 2010). Contrary to necrosis, apoptosis is a highly regulated process of cell death (Green, 2011). There is no specific marker for necrotic tissue regions and currently the assessment is based on histological evaluation of H&E-stained tissue (Robbins et al., 2010).

Among cancers, necrosis has an important role in histological classification and is generally associated with poor prognosis. For

instance, in lung (Swinson et al., 2002), colorectal (Pollheimer et al., 2010), and thyroid carcinoma (Caruso et al., 2011), tumor necrosis has been shown to correlate with shorter survival times. Generally, the presence of necrosis is a sign of aggressive disease and is a result of local hypoxia within a tumor (Hockel & Vaupel, 2001). However, in some tumors, necrosis can be also an indication of patient response to neoadjuvant therapy (Vaynrub et al., 2015).

In preclinical cancer research on tumor models, tumor necrosis is commonly used as a metric of treatment effectiveness when investigating anticancer agents. Furthermore, tumor necrosis may serve as a metric of quality of archived specimens in biobanks (Muley et al., 2012).

2.2.2 Tumor-infiltrating lymphocytes

Accumulating evidence suggests that the host immune system may have a key role in combatting cancer cells through anti-tumor immunity (Luen et al., 2017). Tumor-infiltrating lymphocytes (TILs) are mononuclear leukocytes that surround and infiltrate tumors and are considered as a potential biomarker of immunogenicity.

TILs are usually composed of a heterogeneous mixture of different leukocyte subtypes that can be identified with IHC (Ruffell et al., 2012). However, histological assessment of the total amount of TILs in H&E-stained tumor specimens is the most common method of detection (Savas et al., 2015).

The abundance of TILs is often associated with a more favorable prognosis in different cancers (Fridman et al., 2017). The first evidence of the positive correlation between a high degree of TILs and favorable prognosis was reported in breast cancer (Sistrunk & Maccarty, 1922). In addition to breast cancer, ample evidence suggests an association of TILs and longer survival in ovarian cancer (Santoiemma & Powell, 2015) and melanoma (Lee & Margolin, 2012). In addition, findings in breast cancer suggest that TILs might be an important marker for selecting patients for immunotherapies (Loi et al., 2014).

2.2.3 Breast cancer outcome prediction

Histological examination has a significant role in prognostication of breast cancer patients. Histological grade and type, expression status of cell receptors, and dissemination of cancer cells to axillary lymph nodes are all based on histologic analysis and are among the most important prognostic factors (Tavassoéli & Devilee, 2003).

In breast cancer, histological grade is a three-level classification of tumor tissue differentiation that considers specific tissue entities (tubular differentiation, nuclear pleomorphism, and mitotic count) (Elston & Ellis, 1991). Grade 1 is the lowest grade level (most similar to healthy tissue) and has the best prognosis, while grade 3 is the highest grade level and is associated with poor prognosis (Rakha et al., 2008).

Breast cancer originates from epithelial tissue and results in morphologically diverse carcinomas with differential survival profiles. The most important histological types of breast tumors include *in situ* carcinomas, invasive ductal carcinoma, invasive lobular carcinoma, and carcinoma of special type (Fritz et al., 2010).

Estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor 2 receptor (HER2) are cell receptors that are regularly assessed by IHC for diagnosis and prognosis. Patients with tumors that express ER, PR hormone receptors usually have a more favorable prognosis (Li et al., 2003). HER2 is a protein that is often overexpressed in aggressive disease but is expressed at only low levels in normal breast tissue (Hoff et al., 2002). Tumors are classified as either negative or positive with regards to the expression status of these receptors. Patients with tumors that are negative for all these three receptors (triple-negative breast cancer) have a poor prognosis (Carey et al., 2006).

Furthermore, the extent of breast cancer spread is assessed by histological examination of axillary lymph nodes. Disease with lymph node involvement is associated with rapid tumor growth and is one of the strongest prognostic factors (Toikkanen & Joensuu, 1990).

2.3 Digital pathology

Digital pathology is an interdisciplinary field at the intersection of pathology and digital technologies (Griffin & Treanor, 2017). Digitization of histological specimens into digital image format is the key component in digital pathology. High-resolution whole-slide scanners enable accurate digitalization of histological specimens with sub-micrometer resolution into whole-slide images (WSIs) (Pantanowitz et al., 2011). During the last two decades, digital pathology has created a new ecosystem around WSIs, which aims to improve conventional pathology workflows. These improved workflows allow for more efficient solutions to manage and share samples and also offer novel opportunities to advance interpretation of the histological specimens.

Digital pathology is still a young field and various naming conventions have been used in the literature. Influenced by digital mammography, early studies used the term *computer-aided diagnosis* (CAD) broadly for digital pathology applications concerning image analysis. The term *telepathology* (practice of pathology at a distance) largely overlaps with modern digital pathology applications that aim for improved sharing of digitized tissue sections. Closely related to telepathology, *virtual microscopy* has also been used for data sharing, educational applications, and WSI management solutions. Furthermore, the term *computational pathology* is frequently used in the literature for computerized analysis applications.

2.3.1 Components of digital pathology

There are five main components in digital pathology, namely digitization, new interface, data sharing, data management, and computerized analysis.

Digitization: A whole-slide scanner takes a glass slide with a prepared specimen as an input and transforms it into digital format (i.e. WSIs) (Pantanowitz et al., 2011). Briefly, a slide scanner is composed of a microscope connected to a light-sensitive sensor and robotics are responsible for moving the glass slides, focusing, and

changing objectives. The resulting WSIs are constructed of several individual images, each capturing a field of view (FOV), which are subsequently stitched together. Depending on the objective used, a modern slide scanner is capable of digitizing a large specimen in only a couple of minutes[1]. Commonly, objectives with 5×, 10×, 20×, and 40× magnification are used.

New interface: Digitization of histological specimens allow for viewing the WSIs on displays and computer screens instead of examining them through the microscope eyepiece. The new interface can result in reduced time requirements and improved ergonomics (Thorstenson et al., 2014; Vodovnik, 2016). Furthermore, WSIs allow for a larger viewing field when compared with traditional microscopes. Although this new way of interacting with the slides differs considerably from the traditional technique, several studies have confirmed the value of digital pathology for diagnostic pathology in routine pathology (Bauer & Slaw, 2014; Snead et al., 2016; Stathonikos et al., 2013; Vodovnik, 2016). However, digitization can produce imaging artefacts such as incomplete scanning and out-of-focus issues and therefore quality verification is required (Al-Janabi et al., 2012).

Data sharing: In addition to the novel interface, digitization enables easy sharing of WSIs. Unlike glass slides, digitized samples can be shared and accessed almost immediately throughout the world via the Internet (Farahani & Pantanowitz, 2015). Example usages of WSI sharing include education, research, remote work, and consultation (Al Habeeb et al., 2012; Rocha et al., 2009). For example, scanned samples can be shared with pathologists who are experts in their subfield for second opinion consultation.

Data management: Another benefit of digital pathology is improved archiving of samples (i.e. data management). Redundant digital storage technologies can be utilized in backing up WSI collections (Bhargava & Madabhushi, 2016). Additionally, querying a sample from a digital database is more convenient when compared with retrieving a slide from a pathology archive. However, high-resolution digitization of specimens will result in WSIs of large size

[1] https://scanner-contest.charite.de/en/results/

whose dimensions can surpass 100 000 pixels. Therefore, scanning large collections of samples will lead to substantial storage requirements (Hamilton et al., 2014).

Computerized analysis: Digitization of glass slides has opened up new opportunities for assessing samples through computerized analysis (Madabhushi & Lee, 2016). The same computer vision algorithms that are successfully applied in solving complex object recognition and image analysis problems can now be integrated into analysis of WSIs. Computer vision analysis of WSIs can facilitate disease diagnosis, for example by automatically detecting entities of interest, such as mitoses (Veta et al., 2015) and immune cells (Janowczyk & Madabhushi, 2016) or even determining tumor grade (Awan et al., 2017).

2.4 Computer vision in analysis of cancer histology

Computer vision refers to the broad field of computational methods that are used to mimic or even supersede humans' ability to process and understand visual information in digital images. Computer-vision methods include algorithms from image processing, image analysis, and machine learning (Klette, 2014).

Computer-vision methods have proven their utility in analysis of multifaceted medical imaging data. For instance, recent studies demonstrated the applicability of computer vision in automated screening for diabetic retinopathy (Gulshan et al., 2016) and classification of skin lesions (Esteva et al., 2017). Rapid technological advances in both data-storage solutions and computational resources have increased the adaptation of digital pathology workflows, resulting into more frequent digitization of histological tissue specimens. This has subsequently led to increased interest towards the adaptation of computer vision and machine-learning methods for analysis of cancer histology.

Histopathological analysis offers a versatile and complex environment for computer-vision solutions. Challenges vary from technical aspects, such as color normalization (Khan et al., 2014), to high-level efforts in mining and linking visual information to patient outcome (Beck et al., 2011). H&E is the principal stain for histology

and thus a large portion of computer-vision analysis, such as in this thesis, is focused on analysis of H&E-stained tissue sections. Nevertheless, a large number of studies has investigated the quantification of IHC (Sheikhzadeh et al., 2018; Tuominen et al., 2010).

Computer-vision applications of tissue specimens stained for H&E can be divided into the following three separate levels depending on the scale of the entity of interest: cell, region, and sample level. Cells are the fundamental building blocks of tissue and therefore computerized analysis of cells and nuclei has been of a major interest (Al-Kofahi et al., 2010; Xu et al., 2016). In particular, detection of proliferating cells has gained considerable attention due to the prognostic role in different cancer types (Veta et al., 2015). Moreover, increased interest and understanding of the role of immune cells has resulted in studies aiming to quantify TILs in tumor samples stained for H&E (Fatakdawala et al., 2010; Janowczyk & Madabhushi, 2016).

Region-level analysis covers another set of fundamental entities of cancer histopathology. Segmentation and classification of benign or cancerous tissue structures have been studied, such as glands in colon (Sirinukunwattana et al., 2017), and prostate tissue (Tabesh et al., 2007), or breast cancer metastases in lymph nodes (Bejnordi et al., 2017), and stromal tissue (Fouad et al., 2017). In the case of stromal tissue, a cellular analysis approach can be challenging when the entity of interest is not composed of cells or cells comprise only a small area of the tissue entity of interest. Therefore, segmentation of a specimen into homogeneous tissue regions with a regional approach may be beneficial. Sliding window or superpixel segmentation are commonly used to divide images into regions for subsequent classification.

Instead of dissecting a tissue specimen into separate entities, sample-level analysis aims to automatically categorize the whole specimen. Nevertheless, both cell-level and region-level analysis can serve as an intermediate step in sample-level analysis. A common example of sample-level analysis is computerized grading. Automated tumor grading has been studied broadly in different cancer types, such as breast (Basavanhally et al., 2013), prostate (Jafari-Khouzani & Soltanian-Zadeh, 2003), and glioma (Ertosun & Rubin, 2015). Although tumor grading is of prognostic value, computerized analysis is not strictly limited to follow grading for prognostication. Large

patient cohorts allow for systematic analysis of morphological features in multi-parametric fashion, which may be used to directly predict patient prognosis without introducing intermediate proxies such as grade. Promising results from such an approach have been demonstrated in lung (Yu et al., 2016) and breast cancer (Beck et al., 2011).

Several thorough summaries of computer-vision applications for analysis of digitized histological specimens have been published (Bhargava & Madabhushi, 2016; J.-M. Chen et al., 2017; Gurcan et al., 2009; Litjens et al., 2017; Robertson et al., 2017; Veta et al., 2014).

2.4.1 Texture analysis

Texture analysis has been a common approach for computerized analysis of histological specimens. Entities present in specimens often lack clear boundaries and homogeneous content (such as objects in regular photographs). Popular texture descriptors in histological analysis include local binary patterns (LBPs) (Pietikäinen et al., 2011), grey-level co-occurrence matrix (GLCM) (Haralick et al., 1973), and Gabor filters (Fogel & Sagi, 1989). A study using texture descriptors proposed a segmentation into epithelial and stromal tissue structures in TMAs of colorectal tumor specimens (Linder et al., 2012). Similarly, texture analysis was used for stroma-epithelium segmentation in breast and ovarian cancer (Signolle et al., 2008, 2010). Another study in colorectal cancer proposed multiclass classification for segmenting specimens into seven different tissue entities and backgrounds (Kather et al., 2016).

2.4.2 Deep learning

During the last 5 years, use of deep learning in computerized analysis of histological specimens has become increasingly popular (Janowczyk & Madabhushi, 2016). This is due to the significant impact deep learning has had in visual object recognition, speech recognition, and in many other data domains (LeCun et al., 2015). Deep learning is a group of machine-learning methods that learn hierarchical data representations in increasing abstraction levels

(Schmidhuber, 2015). Recently, deep learning was adapted for detection of breast cancer cells in axillary lymph nodes (Bejnordi et al., 2017). In addition, the feasibility of deep learning has been demonstrated broadly in different tasks, including cell detection and classification (Cireşan et al., 2013; H. Wang et al., 2014); in regional analysis such as segmentation of epithelial tissue (H. Chen et al., 2017; Xu et al., 2016); and in tumor grading (Ertosun & Rubin, 2015).

3 AIMS OF THE STUDY

The overall aim of this doctoral thesis was to investigate the utility of computer vision in characterization of tumor tissue and outcome prediction through analysis of digitized H&E-stained specimens.

Specifically, the aims were to:

1. Develop a method for quantification of tumor viability in lung cancer xenografts.

2. Develop a method for quantification of infiltrating immune cells in breast cancer.

3. Study computerized patient outcome prediction in breast cancer.

4 MATERIALS AND METHODS

4.1 Study specimens

4.1.1 Lung cancer xenograft WSI cohort (I)

In Study I, we investigated tumor viability assessment in a cohort of 72 tumor sections of human non-small cell lung cancer (NSCLC) mouse xenografts. Human NSCLC adenocarcinoma cells (NCIH460-LNM3512) were implanted subcutaneously into mice. Once the largest tumor diameter reached 19 mm in length, the mice were sacrificed and the primary tumors were excised, cut into halves and fixed with 4% paraformaldehyde. The paraffin-embedded tumor tissues were cut into sections of 5 to 7 μm and then stained with H&E. A total of 72 WSIs were scanned. After an image quality check, a subset of 56 WSIs with minimal out-of-focus areas were chosen for further analysis.

The mice were maintained in the Meilahti Experimental Animal Center according to Institutional Animal Care and Use Committee of the University of Helsinki and Institutional Review Board guidelines. The study protocol was approved by The National Animal Experiment Board of Finland (permit number ESAVI/6492/04.10.03/2012).

4.1.2 Breast cancer WSI cohort (II)

In Study II, FFPE tumor samples from 20 breast cancer patients were used to investigate computerized quantification of infiltrating immune cells. The patients (**Table 1**) were operated for primary breast cancer within the Hospital District of Helsinki and Uusimaa, Finland. The samples were anonymized and all patient-related data and unique identifiers were removed. Therefore, the study did not require ethical approval in compliance with Finnish legislation regulating human tissues obtained for diagnostic purposes (act on the use of human organs and tissue for medical purposes 2.2.2001/101). The Head of the Division of Pathology and Genetics approved of the use of the samples. From each FFPE block, two 3.5-

μm thick consecutive sections were cut and stained with H&E and for CD45.

Table 1. Patient characteristics of the breast cancer WSI cohort

Patient characteristics	N	%
Histological type		
Ductal carcinoma	13	65
Lobular carcinoma	3	15
Medullary carcinoma	2	10
Adenosquamous carcinoma	1	5
Histological grade		
Grade I	3	15
Grade II	3	15
Grade III	14	70

4.1.3 Breast cancer TMA cohort (III)

For Study III, we pooled two breast cancer patient cohorts with TMA samples and the available follow-up information. For the first dataset, we identified 2 864 women diagnosed with breast cancer in 1991 and 1992 using the Finnish Cancer Registry files. The cohort (FinProg Breast Cancer Database) is accessible online[2]. The other cohort comprises tissue samples and follow-up information from 527 women with invasive ductal breast cancer treated at the Department of Surgery and Oncology, Helsinki University Hospital, between January 1987 and December 1990. Clinical and pathological information associated with the patients were extracted from the hospital and laboratory records.

From this pooled patient cohort, we excluded patients with lobular or ductal carcinoma in situ, synchronous or metachronous bilateral breast cancer, other malignancies (except for basal cell carcinoma or cervical carcinoma in situ), distant metastasis, and those who did not undergo breast surgery. We included only those patients who had specific survival information available, those with available breast

[2] http://www.finprog.org/

cancer tissue samples, and those who had a digitized TMA spot image where the area of tissue was greater than 400 000 pixels. Altogether this yielded 1 299 patients with associated TMA samples, clinical characteristics, and follow-up information. The patients were randomly divided into a separate training set (66%) and test set (33%) (**Table 2**). The median follow-up of patients in the patient cohort alive at the end of follow-up period was 15.9 years (range, 15.0-20.9 years).

Project-specific ethical approval for the use of clinical samples and retrieval of clinical data was approved by the local operating ethics committee of The Hospital District of Helsinki and Uusimaa (DNo 94/13/03/02/2012). Approval was also obtained from the National Supervisory Authority for Welfare and Health (Valvira) for the use of human tissues for research (7717/06.01.03.01/2015).

Table 2. Patient characteristics of the training and test sets in the breast cancer TMA cohort

Variables	Training set (N=868)		Test set (N=431)		P-value
	%	N	%	N	
Number of positive lymph nodes					
mean	1.4		1.2		0.407
0	58	504	59	253	
1-3	24	206	23	99	0.323
4-9	8	73	9	38	
>10	3	30	2	7	
Unknown	6	55	8	34	
Tumor size, per mm					
mean	23.7		23.2		0.817
Unknown	3	28	5	22	
Histological grade					
Grade I	16	143	19	83	
Grade II	34	296	36	154	0.086
Grade III	23	197	18	76	
Unknown	27	232	27	118	
Histological type					
Ductal	76	662	77	333	0.742
Lobular/Special	24	206	23	98	
Age, years					
≤39	7	63	7	30	
40-49	21	186	24	103	
50-59	27	234	22	94	0.353
60-69	20	172	21	91	
≥70	25	213	26	113	
ER					
Negative	29	248	27	116	0.572
Positive	62	538	64	274	
Unknown	9	82	10	41	
PR					
Negative	42	362	41	177	0.803
Positive	49	423	50	215	
Unknown	10	83	9	39	
HER2					
Negative	72	623	74	321	0.713
Positive	17	146	16	70	
Unknown	11	99	9	40	

4.2 Sample digitization

All the tumor tissue samples used in this thesis were digitized with an automated whole-slide scanner (Pannoramic 250 FLASH, 3DHISTECH, Budapest, Hungary). The scanning was performed with a Plan-Apochromat 20× objective (numerical aperture 0.8) and a VCC-F52U25CL camera (CIS, Tokyo, Japan) equipped with three (1 224 × 1 624 pixels) charge-coupled device sensors. The pixel size of the sensors is 4.4 × 4.4 μm. In combination with the 20× objective and a 1.0 adapter, the image resolution was 0.22 μm/pixels. Images were compressed to wavelet file format (Enhanced Compressed Wavelet, ECW, ER Mapper, Intergraph, Atlanta, Georgia, USA) with a compression ratio of 1:9. The compressed virtual slides were uploaded to a WSI management server (WebMicroscope, Fimmic, Helsinki, Finland).

4.3 Image annotation

In Study I, we annotated 671 single tissue entity images (STEIs) for training (N=177) and for testing (N=494) of the tissue entity classifier. The STEIs (945 × 945 pixels) were cropped from homogeneous tissue regions, representing only one of the tissue entities of interest (viable tumor, necrotic tumor, or host tissue). The training set STEIs were extracted from four WSIs and the test STEIs were extracted from 23 WSIs. Furthermore, we manually annotated viable and necrotic tumor tissue regions in each of the 52 WSIs that were not used in extraction of the training STEIs. An online WSI-management software (WebMicroscope, Fimmic, Helsinki, Finland) was used in annotating the STEIs. A raster graphics editor (Adobe Photoshop CS6, Adobe Systems, Mountain View, California, USA) was used in annotation of the WSIs.

In Study II, we annotated a training set of image regions of various size (N=1 116) from 20 WSIs. Four different tissue entities (leukocyte-rich, epithelial, stromal, and adipose) and background were considered. The manual annotation of the H&E-stained WSIs was guided with paired and CD45-stained WSIs. Leukocyte-rich tissue

regions were identified with the IHC marker. The IHC staining guided the selection of the other tissue entities into regions that were negative for CD45 expression and therefore did not contain immune cells. The training set was annotated with a raster graphics editor (Adobe Photoshop CS6, Adobe Systems, Mountain View, California, USA). Moreover, we randomly selected 200 images (1 000 × 1 000 pixels) from the 20 WSIs (10 random images per WSI). For ground truth, three pathologists assessed the proportional amount of each tissue entity of interest within the test images.

In Study III, training of the outcome-prediction model was guided with follow-up information and therefore no training data were annotated. However, for comparing the model with human experts, the test-set TMAs (N=431) were examined by three pathologists and given a visual risk score. The visual risk score (low or high) is a pathologist's assessment of a patient's risk based on the visual features present in the TMAs. Additionally, one pathologist annotated the following tissue entities in the test TMAs: mitoses (0 vs. 1 vs. >1), pleomorphism (minimal vs. moderate vs. marked), tubules (≤10 vs. 10-75 vs. >75%), necrosis (absent vs. present), and quantity of TILs (low vs. high). All annotations in Study III were performed with an online WSI-management software (WebMicroscope, Fimmic, Helsinki, Finland).

4.4 Computer-vision methods

4.4.1 Texture descriptors (I, II)

A texture descriptor defined as a joint distribution of the local binary pattern (LBP) and the rotation invariant variance (VAR) descriptor was applied in studies I and II (Ojala et al., 2002; Pietikäinen et al., 2011). Prior to feature extraction, input images were converted into grayscale with following channel wise weights: 0.2989, 0.5870, and 0.1140. In the case of LBP, only the *rotation invariant 2-uniform* (i.e. *riu2* descriptors) was considered. Both the LBP and VAR descriptors are parametrized with the pattern radius (r) and number of sampling points (p). In Study I, two joint distributions of LBP and VAR were

extracted with (p,r)-parameter pairs of (3,8) and (4,16). For classification, the feature vectors were concatenated together. In Study II, only one descriptor (4,16) was considered. MATLAB implementations for the texture descriptors (available online[3]) were used.

4.4.2 Image description with a deep CNN (II, III)

Image descriptors extracted with deep convolutional neural networks (CNNs), pre-trained with the ImageNet (Jia Deng et al., 2009) database of natural images, were utilized in discrimination of tissue entities of interest. In Study II, we exploited the VGG-F (Chatfield et al., 2014) by reading the fully connected activations from the network's penultimate layer. Superpixels (Achanta et al., 2012) scaled to match the input of the network (224 × 224 pixels) served as an input for the CNN, resulting in a descriptor of 4 096 bins.

We employed the VGG-16 (Simonyan & Zisserman, 2014) network for feature extraction in Study III. Instead of reading the fully connected activations, we took advantage of the last convolutional layer of the CNN. This allowed us to input an image of arbitrary size into the network, resulting in an activation tensor of 512 channels and row and column number being dependent on the input image size. Applying first principal component analysis (PCA) to compress the local activation, we aggregated the descriptor into a 1-dimensional vector with improved Fisher vector (IFV) encoding.

The mean of the ImageNet training images was in normalization of the intensity values of input images. A MATLAB toolbox (Vedaldi & Lenc, 2014) for implementation of CNNs was used. The pre-trained CNNs are available for download online[4].

4.4.3 Homogenous kernel maps (I, II)

Homogenous kernel maps (Vedaldi & Zisserman, 2012) were utilized in Studies I and II together with a linear support vector machine

[3] http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab/

[4] http://www.vlfeat.org/matconvnet/pretrained/

(SVM) classifier. Kernel maps facilitate the use of non-linear kernels in large-scale classification problems by approximating kernel functions. This in turn enables the use of linear SVM that are rapid to train and test and simultaneously enable the use of a more flexible model. The feature map offers a low-dimensional approximation for many popular kernels (such as intersection and chi-square kernels) used in computer vision. Studies I and II applied the chi-square feature map for texture descriptors. A computer-vision toolbox (Vedaldi & Fulkerson, 2010) for MATLAB offered an implementation for the kernel map.

4.4.4 Improved Fisher Vector encoding (III)

Fisher vector (FV) encoding is a method for orderless feature pooling (Perronnin & Dance, 2007). A feature pooling encoder takes local image descriptors as an input and constructs a single output for further analysis (such as for classification). The pooling encoders that do not maintain the spatial relationship of the local image descriptors are considered orderless encoders. FV exploits a Gaussian Mixture Model (GMM) as an intermediate quantizer and describes the local image descriptors with the mean and the covariance of the soft assignments of GMM. The IFV encoding further introduces the use of signed square rooting and L^2 normalization for improved classification performance (Perronnin et al., 2010). A MATLAB implementation provided in a toolbox for computer vision (Vedaldi & Fulkerson, 2010) was applied for computation of IFV and GMM.

4.4.5 Linear support vector machine (I, II, III)

SVMs are a group of supervised learning methods for classification and regression (Cortes & Vapnik, 1995). Briefly, a SVM is defined as a maximum margin classifier, or a classifier that constructs a hyperplane in the feature space that separates two categories by the largest margin. SVM is a linear classifier by nature. However, incorporation of nonlinear kernel tricks (Theodoridis & Koutroumbas, 2009) that transform the feature space allow for design of a nonlinear SVM. In this thesis, only linear SVM was utilized. A MATLAB

implementation provided in a toolbox for computer vision (Fan et al., 2008) was applied.

4.5 Statistical analysis

Classification results were evaluated with F-score, area under receiver operating characteristics curve (AUROC), and with accuracy, sensitivity, specificity, and precision. Cohen's kappa value (κ) and Pearson's product-moment correlation were used for evaluation of agreement. The Kaplan-Meier method was used in the analysis of the survival profiles (Kaplan & Meier, 1958) and the log-rank test was used in comparison of the profiles. The Cox proportional hazard model (Cox, 1972) was utilized to estimate the effect size (hazard ratio, [HR]) and to adjust for covariates. C-statistics (concordance) were used to compare the discriminative accuracy of survival models (Gönen & Heller, 2005). The chi-squared test and the Kruskal-Wallis test were used in comparison of categorical and continuous variables, respectively. Statistical tests with $P<0.05$ were considered statistically significant. Statistical analyses were performed with R and MATLAB programming languages.

5 RESULTS

5.1 Assessment of tumor viability

A computational method that utilized texture analysis was developed for quantification of tumor viability in WSIs of H&E-stained NSCLS xenograft tumor samples (**Figure 1**). To quantify tumor viability, the WSIs were segmented into the following three distinct tissue entities: non-viable (i.e. necrotic) tumor tissue, viable tumor tissue regions, and host tissue comprising mostly stromal, adipose, or muscle tissue. Separation of these main tissue regions facilitated tumor viability assessment.

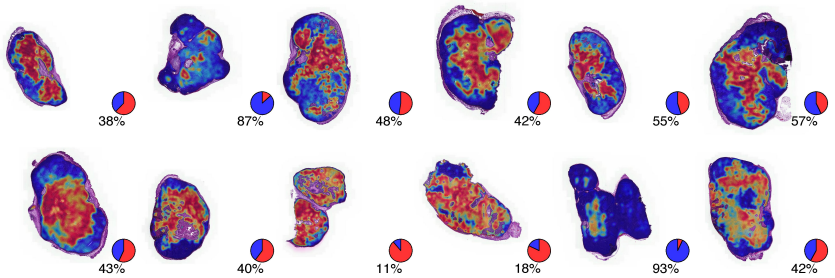

Figure 1. Examples of 12 whole slide images (WSIs) analyzed for viability. Heat map displays the predicted viability superimposed on top of the hematoxylin and eosin (H&E)-stained tissue specimens. Red color indicates that a tissue region is classified as necrotic and blue indicates viable tissue. Pie charts show the ratio of viable tissue to whole tumor region. Adapted from (Turkki et al., 2015).

We hypothesized that the tumor samples stained only for basic morphology could be characterized with algorithms proven to perform well in analysis of textures. Therefore, a feature combining LBP and VAR texture descriptors was considered.

The large size of the WSIs requires division of the images into smaller image batches for analysis. Tiling the WSIs into sub-images (3 968 × 3 968 pixels) and analyzing the sub-images with a sliding window classifier enabled the processing of the gigapixel-sized WSIs. A sliding window of 128 × 128 pixels with displacement of 64 pixels was used together with a feature mapping and SVM classifier to produce a segmentation map of each NSCLC tumor sample.

5.1.1 Human expert guided training

A training set of STEIs was created to produce a collection of examples representing the three tissue entities. The aim of this approach was to eliminate the use of unclear tissue regions (i.e. regions containing several tissue entities of interest in a single image) from training. Our experience and hypothesis were that using clean training data would result in more robust classification. In total, 177 STEIs were labeled for training, of which 57, 52, and 68 represented viable tumor, necrotic tumor, and non-tumorous host tissue regions, respectively. Using the sliding window approach, the training STEIs were processed for extraction of the texture descriptors, which were used to train a linear SVM classifier. The classifier cost parameter was selected via a three-fold cross-validation parameter sweep in the training set.

5.1.2 Comparison with human experts

We compared the performance of the suggested approach to those of human experts in discrimination of viable and non-viable tumor regions in a separate test set of 494 STEIs (N=242, viable tumor; N=252, non-viable tumor). An agreement of 95% with a ROCAUC of 0.995 was obtained. In discrimination between viable and necrotic tumors, 23 human expert-labeled viable STEIs were misclassified, whereas only two non-viable STEIs misclassified. This corresponds to an agreement of κ=0.90 (95% CI 0.86–0.97) and a sensitivity and specificity of 91% and 99%, respectively.

5.1.3 Evaluation on WSIs

We next evaluated the computerized tumor viability assessment in 52 NSCLC WSIs that were annotated by human experts. At the sample level, a correlation of r=0.79 (95%CI 0.66–0.87; P<0.0001) was obtained. At the pixel level, the average agreement between computerized assessment and human expert assessment was 83.3%.

5.2 Quantification of infiltrating immune cells

A method utilizing specific antibody staining in training data labeling and a deep CNN in feature extraction was developed for quantifying the degree of immune cell infiltration in WSIs of H&E-stained breast cancer samples (**Figure 2**). The computational pipeline adopts a transfer learning in the analysis of digitized histological samples through a pipeline that comprises superpixel segmentation, feature extraction with a deep CNN, classification, and post-processing. The WSIs were tiled (3 000 × 3 000 pixels) for analysis.

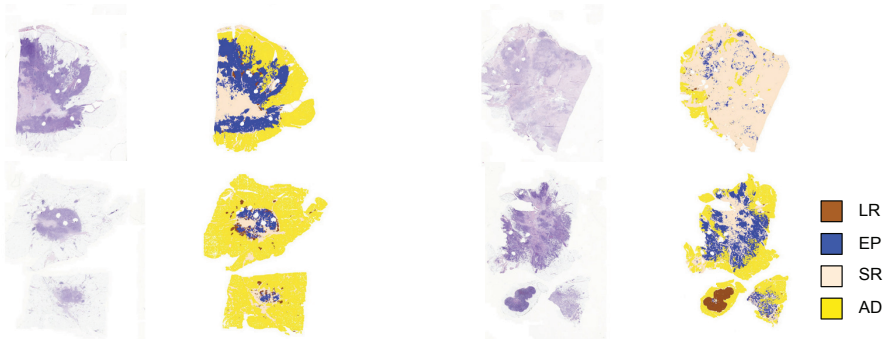

Figure 2. Examples of hematoxylin and eosin (H&E)-stained breast tumor specimens segmented into the following five categories: leukocyte rich (LR), epithelial (EP), stromal (SR), adipose tissue (AD), and background (BG). Adapted from (Turkki et al., 2016a).

5.2.1 Protein expression guided training

Objective labeling, or collection of the ground truth, is challenging due to the complex nature of histological specimens. We took advantage of two consecutively cut tumor sections, staining one section with the pan-leukocyte CD45 marker and the other with H&E. The annotation of training examples in the WSIs of H&E-stained tumor samples was guided with the specific signal present in the digitized IHC samples.

In total, we collected 1 116 separate tissue regions from 20 WSIs representing immune cell-rich and -poor regions. Five different tissue entities were considered, namely TIL-rich tissue regions (LR), epithelial tissue with none or few TILs (EP), stromal tissue with none or few TILs (SR), adipose tissue with none or few TILs (AD), and background (BG). Guiding the annotation process with IHC allowed

us to identify smaller clusters of TILs within the large WSIs that would have been difficult to identify otherwise. Similarly, in annotation of TIL-poor regions we could easily verify the absence of TILs.

The annotated regions were divided into superpixels for training, serving as an input of the classifier after scaling to 244 × 244 pixels. The antibody-guided annotation resulted in a total of 123 442 superpixels that represent different tissue entities of interest. Three-fold cross-validation in the training data was used for optimizing the cost parameter of the classifier.

5.2.2 Image descriptor comparison

We studied the suitability of a deep CNN to describe and discriminate the different tissue categories. By performing 10 random three-fold cross-validation rounds, we compared features extracted with the deep CNN to texture descriptors. The results showed that the fully connected activations extracted from the penultimate layer of the VGG-F network provided stronger discrimination than the texture descriptors based on LBP and VAR. The overall F-score for the transfer-learning approach was 0.96 whereas with texture descriptors the F-score was 0.92 (**Table 3**). Furthermore, the method reached a sensitivity of 91% (range, 88%–92%), specificity of 100% (range, 100%–100%), and a precision of 96% (range, 96%–97%) to discriminate TIL-rich and TIL-poor superpixels.

Table 3. Discrimination of tissue entities according image descriptor

Descriptor	Mean F-score (range)					
	LR	EP	SR	AD	BG	Overall
LBP/VAR	0.87 (0.86-0.88)	0.87 (0.85-0.88)	0.85 (0.84-0.87)	0.92 (0.91-0.92)	0.95 (0.95-0.96)	0.89 (0.84-0.96)
LBP/VAR-KHCI2	0.88 (0.87-0.89)	0.90 (0.88-0.90)	0.89 (0.87-0.89)	0.94 (0.94-0.95)	0.97 (0.97-0.97)	0.92 (0.87-0.97)
VGG-F	0.94 (0.92-0.94)	0.96 (0.96-0.96)	0.96 (0.95-0.96)	0.98 (0.97-0.98)	0.99 (0.99-0.99)	0.96 (0.92-0.99)

LBP/VAR, local binary pattern and local variance descriptors; LBP/VAR-KCHI2, local binary pattern and local variance descriptors with chi-square kernel map; VGG-F, local image descriptors extracted with the VGG-F network. Tissue entities of interest: LR, leukocyte rich; EP, epithelium; SR, stroma; AD, adipose; BG, background

5.2.3 Comparison with pathologists

Using a leave-one-out strategy, we analyzed all 20 WSIs. Comparison of the TIL assessments from two pathologists with the computerized assessment showed an agreement of 90% (κ=0.79). Inter-agreement of 90% (κ=0.78) was observed between the two pathologists, which is on par with the computerized assessment.

Detailed analysis revealed a clear pattern in the pathologists' assessments that favored numbers that are divisible with 5% in evaluation of TIL percentage. Naturally, computerized methods do not have similar bias. The greatest differences between the computerized assessment and pathologists' visual assessment were seen in the range between 25% to 75%. Interestingly, this phenomenon was also observed between the pathologists.

Correlation analysis indicated the largest disagreement in TIL quantification when compared to the other tissue categories, suggesting this to be the most difficult to quantify. Analysis showed a high correlation (r>0.90) in assessment of TIL-poor tissue entities, while the correlation was more moderate in assessment of TILs. On average, the correlation between the pathologists and the computerized methods was r=0.66, while the pathologists' assessments had correlations of r=0.82.

5.3 Patient outcome prediction

We developed a computerized pipeline that takes a digitized TMA spot image as an input and classifies it into a low or high digital risk score (DRS) group (**Figure 4**). The risk grouping is learned in a training set of images of H&E-stained TMAs using image descriptors extracted with a deep CNN.

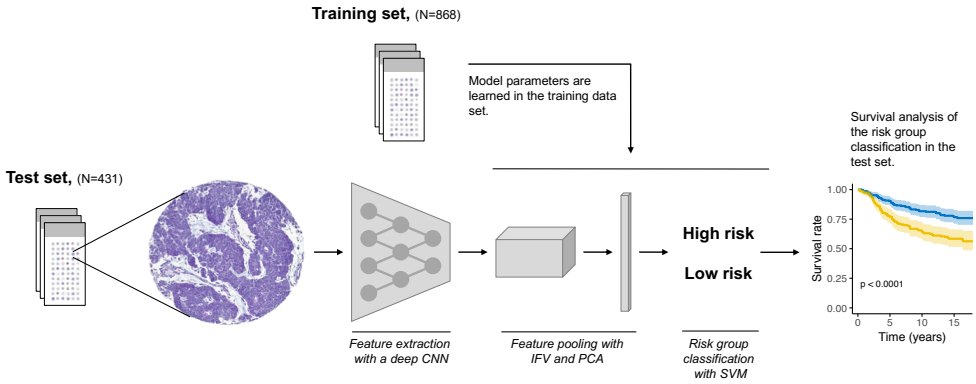

Figure 3. Method for patient outcome prediction in tumor tissue images.

5.3.1 Survival status guided training

We first divided the training set into low-risk and high-risk groups based on patient follow up. Those patients who died of breast cancer within 10 years after diagnosis were considered as examples of high-risk cases. The remaining patients (i.e. those who did not die of breast cancer during the follow-up time or within 10 years) were labeled as examples of low-risk patients. Utilizing deep-CNN activations and feature aggregation, each training set sample was then captured into one feature vector and used together with the risk label to train a linear SVM classifier. The SVM classified the samples into a low or high DRS group. In total, our training set comprised 868 tumor tissue images.

5.3.2 Associations with clinicopathological variables

With the DRS classifier, we analyzed the test set of 431 patients with tumor tissue images. The analysis of the DRS grouping revealed significant differences in clinicopathological variables (**Table 4**). Patients who were classified into the low DRS group more often had lower grade tumors (P=0.014), smaller tumors (P<0.001), and less frequently had positive lymph nodes (P=0.003). These tumors were also more often negative for PR when compared with the patients in the high DRS group.

Table 4. Patient characteristics between test set patients classified into low and high digital risk score (DRS) groups

Variables	Low DRS (N=237)		High DRS (N=194)		*P*-value
	%	N	%	N	
Number of positive lymph nodes					
Mean	0.9		1.6		0.003
0	63	150	53	103	0.057
1-3	23	54	23	45	
4-9	6	15	12	23	
>10	1	2	3	5	
Unknown	7	16	9	18	
Tumor size, per mm					
Mean	21.5		25.3		<0.001
Unknown	5	13	5	9	
Histological grade					
Grade I	23	54	22	43	0.014
Grade II	32	75	41	79	
Grade III	14	33	22	43	
Unknown	32	75	22	43	
Histological type					
Ductal	74	175	81	158	0.079
Lobular/Special	26	62	19	36	
Age					
≤39	9	21	5	9	0.140
40-49	27	64	20	39	
50-59	21	49	23	45	
60-69	20	47	23	44	
≥70	24	56	29	57	
ER					
Negative	25	60	29	56	0.443
Positive	65	155	61	119	
Unknown	9	22	10	19	
PR					
Negative	36	86	47	91	0.015
Positive	56	132	43	83	
Unknown	8	19	10	20	
HER2					
Negative	76	181	72	140	0.136
Positive	14	32	20	38	
Unknown	10	24	8	16	

5.3.3 Survival analysis

In survival analysis, women in the low DRS group were found to have both better disease-specific survival (DSS) (P<0.001) and overall survival (OS) (P=0.003) when compared with the high DRS group patients (**Figure 5**). The analysis further indicated a 10-year DSS status of 82% (95% CI 78%-87%) for the low DRS patients. In contrast, the corresponding value in the high DRS group was only 65% (95% CI 58%-73%).

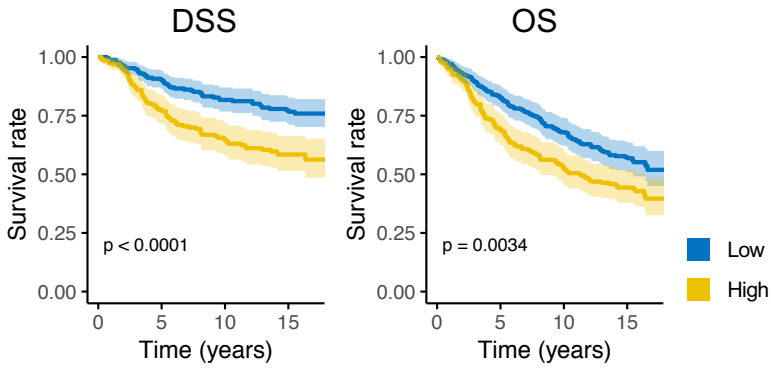

Figure 4. Disease-specific survival (DSS) and overall survival (OS) according to low and high digital risk score (DRS) groups.

We studied the DRS group stratification separately in different patient subgroups with subset analysis. Division according to histological grade displayed significant differences only among grade-I patients (P<0.001) (**Figure 6**). When divided according to hormone receptor status, we observed significant differences in the survival profiles of the DRS groups among ER+ (P=0.025), ER- (P<0.001), and PR- (P=0.003) patients. Similarly, DRS stratification was significant in both HER2+ (P<0.001) and HER2- (P=0.015) patients.

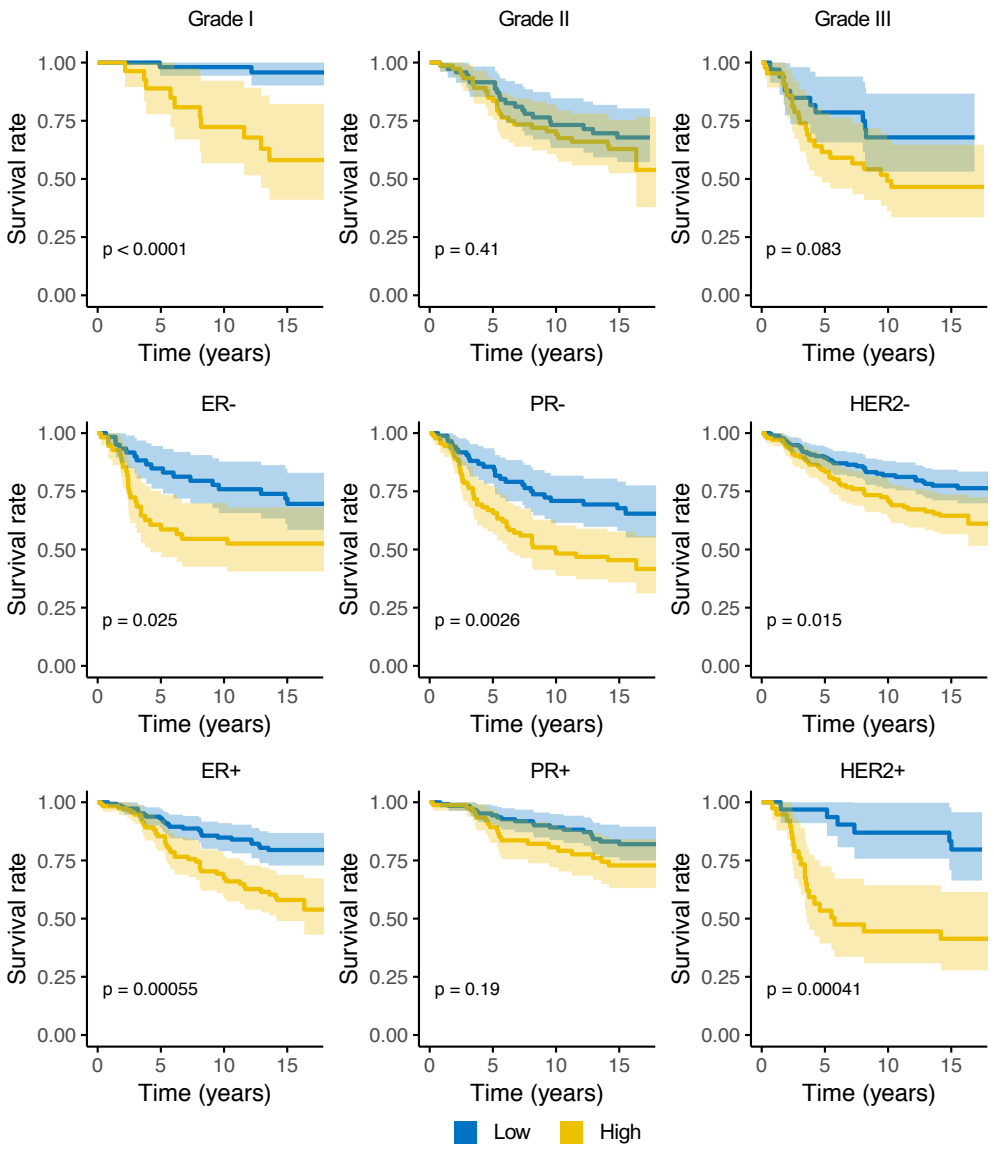

Figure 5. Disease-specific survival according low and high digital risk score (DRS) groups in patients with tumors of different grades and estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) status.

With multivariate survival analysis, we analyzed the effect of the DRS grouping adjusted for other clinicopathological variables. The analysis indicated that the DRS is an independent predictor for DSS (P=0.005) with a HR of 2.06 (95% CI 1.24-3.42) (**Table 5**). The

number of positive lymph nodes (P<0.001) and tumor size (P<0.001) were also identified as independent predictive factors in the multivariate analysis.

Table 5. Multivariate Cox regression

Variables	HR	95% CI	*P*-value
DRS group			
Low Risk	ref		
High Risk	2.06	(1.24-3.42)	0.005
Number of positive lymph nodes			
0	ref		
1-3	1.65	(0.91-2.95)	0.102
4-9	1.47	(0.75-2.85)	0.260
>10	3.45	(1.04-2.85)	0.043
Tumor size			
per mm	1.04	(1.02-1.06)	<0.001
Histological grade			
I	ref		
II or III	1.54	(0.76-3.14)	0.235
Histological type			
Ductal	ref		
Lobular/Special	0.91	(0.42-1.98)	0.808
Age			
≤39	ref		
40-49	0.48	(0.19-1.21)	0.118
50-59	0.49	(0.19-1.23)	0.141
60-69	0.67	(0.27-1.64)	0.377
≥70	1.33	(0.54-3.29)	0.535
PR			
Negative	ref		
Positive	0.41	(0.24-0.69)	<0.001
HER2			
Negative	ref		
Positive	1.16	(0.65-2.07)	0.626

5.3.4 Comparison with visual risk score

Test TMA spot images that were marked by at least one of the pathologists as not evaluable (N=109) were excluded from the analyses. The percent agreement of the pathologists' visual risk scoring in the remaining subset of 322 TMA spot images was 32%. The assessments between pathologists 1 and 3 were significant (P<0.001), whereas the assessments between pathologists 1 and 2 and between 2 and 3 were not significant.

In univariate survival analysis, the merged visual risk score (based on majority vote) was found to be a significant predictor of DSS with a HR of 1.74 (95% CI 1.16-2.61; P=0.006) and C-statistics of 0.58 (95% CI 0.53-0.63) and of DRS with a HR of 2.10 (95% CI 1.40-3.18; P<0.001) and C-statistics of 0.60 (95% CI 0.55-0.65). Interestingly, the survival models resulted in significantly different predictions (P<0.001). Multivariate survival analysis further indicated that the visual risk score and the DRS were both independent predictors for DSS (HR=2.05; P<0.001 for DRS and HR=1.68; P=0.012 for visual risk score), reaching a C-statistics of 0.64 (95% CI 0.58-0.69). An analysis with the morphological cancer features and with the risk scores showed that DRS significantly associated with nuclear pleomorphism and tubule formation, whereas the visual risk score had a significant association with mitotic count, presence of necrosis, and TILs (**Table 6**).

Table 6. Association of morphological cancer features and digital risk score and visual risk score

	Digital risk score					Visual risk score				
Variables	**Low (N=177)**		**High (N=145)**			**Low (N=193)**		**High (N=129)**		
	%	**N**	**%**	**N**	***P*-value**	**%**	**N**	**%**	**N**	***P*-value**
Mitoses										
0	84	148	74	107		96	186	53	69	
1	9	16	15	22	0.095	2	3	27	35	<0.001
>1	7	13	11	16		2	4	19	25	
Pleomorphism										
Minimal	22	39	6	8		21	41	5	6	
Moderate	59	105	60	87	<0.001	72	139	41	53	<0.001
Marked	19	33	34	50		7	13	54	70	
Tubules, %										
<10	81	144	88	127		76	146	97	125	
10-75	14	24	12	18	0.020	20	38	3	4	<0.001
>75	5	9	0	0		5	9	0	0	
Necrosis										
Absent	95	169	98	142	0.370	99	192	92	119	0.001
Present	5	8	2	3		1	1	8	10	
TILs										
Low	90	160	83	120	0.063	91	175	81	105	0.024
High	10	17	17	25		9	18	19	24	

6 DISCUSSION

For over a century, examination of histological tissue sections for detection and diagnosis of cancer has relied solely on the visual interpretation of experienced pathologists. Empowered by digital whole-slide scanners and advanced machine-learning algorithms, histological diagnostics is undergoing a paradigm shift towards precision histology (Djuric et al., 2017). These novel digital pathology workflows are likely to improve both accuracy and throughput of cancer histology assessment and to support personalized cancer care through improved reproducibility and stratification.

The overall goal of this doctoral thesis was to investigate the use of computer vision for characterization and outcome prediction in cancer by mining biologically relevant signals in tumor tissue specimens. Although cancer cells are the foundation of malignant neoplastic diseases, cancer biology cannot be thoroughly understood without considering the complex and dynamic interaction of cancer cells and the tumor microenvironment (Hanahan & Weinberg, 2011). The characteristics of the tumor microenvironment influence patient response to treatments (Tredan et al., 2007) and the ability of the tumor to grow and obtain nutrients (Whiteside, 2008) and the ability of the cancer cells to invade distant organs (Mlecnik et al., 2016). To gain a better understanding of tumor characteristics and the varied features of its microenvironment, novel tools are needed that enable more systematic and accurate quantification and characterization of the composition of tissue specimens (Letai, 2017).

In Study I, we aimed to detect and quantify tumor necrosis in WSIs of H&E-stained human lung cancer xenograft tumors. The quantification of necrosis has not been studied extensively. However, early studies have been published for glioblastoma (Le et al., 2012), kidney carcinoma (Nayak et al., 2013), and liver (Homeyer et al., 2013) tissue images. It should be noted that direct comparison between studies is difficult due to varying experimental setups and different cancer types. However, a more recent study that also used texture analysis reported an accuracy of 85% in detecting necrosis in gastric adenocarcinoma specimens (Sharma et al., 2015). The authors

showed that use of CNN improved necrosis detection accuracy from 73% to 82% when compared to texture descriptors (Sharma et al., 2017). The results obtained in Study I indicated an accuracy of 95% for the proposed texture analysis approach.

Taken together, we demonstrated in Study I that computerized analysis of tumor tissue texture can facilitate accurate quantification of necrosis. Moreover, our experiments showed that the texture features used offer comparable or even superior discrimination when compared with the other methods presented in the literature. Although the material was from an animal model, there is no reason why the method could not be adjusted for human tissue specimens. In fact, we later successfully adapted a similar approach for analysis of human tumor specimens (Turkki et al., 2016b). Although texture analysis can facilitate accurate segmentation of histological specimens into entities of interest (such as necrosis), more recent findings by our group (Turkki et al., 2016a) and others (Bejnordi et al., 2017) suggest that CNNs can yield stronger discrimination.

Recent findings in breast cancer suggest that patients with strong immune infiltration have less aggressive disease (Savas et al., 2015). Particularly in patients with HER2-positive and the triple-negative subtype of breast cancer, the abundance of TILs is linked with a favorable prognosis. Presently, the assessment of TILs is based on visual examination of histological tumor samples stained for H&E (Salgado et al., 2014). In Study II, we investigated the use of computer vision to quantify TILs in WSI of breast tumors. We achieved an agreement of 90% (κ=0.79) between computerized and pathologist assessments. Further analysis showed that this was consistent with the pathologists' inter-observer agreement (90%, κ=0.78).

To expand the texture analysis presented in Study I, we incorporated deep learning into our analysis via a transfer-learning approach. Transfer learning has been shown to improve performance in both texture analysis and object-recognition tasks (Cimpoi et al., 2016). The findings in Study II showed that transfer learning can lead to significant performance improvement in analysis of histological specimens. Although LBP-based descriptors reached high accuracy in TIL quantification (F-score=0.88), the local image descriptors extracted with the VGG-F network reached superior performance (F-

score=0.94). The performance gain was strongest with TILs when compared with other studied tissue entities (epithelium, stroma, adipose, and background). This finding most likely reflects the overall difficulty of discriminating TILs from the other tissue entities of interest.

A large majority of earlier studies, as well as the following, have approached TIL quantification through cell detection and subsequent classification. A challenge for the cell-level approach is the requirement of ground truth that is usually difficult to obtain. This may be the reason why some of the studies have been conducted on limited data sets (Fatakdawala et al., 2010; Nawaz et al., 2015; Panagiotakis et al., 2010). Deep learning has also been adapted for immune cell detection (Janowczyk & Madabhushi, 2016). In that study, a detection accuracy of F-score=0.90 was obtained across 3 064 immune cells. In contrast to the earlier studies, our method was implemented as a region-level analysis; the WSIs were divided into superpixels that were then classified into different tissue categories. However, this limits the analysis into quantification of tissue regions that are densely packed with TILs. Consequently, single immune cells that are scarcely scattered around the tissue compartments are not captured by our method. Despite the region-level approach, our results indicated high agreement with pathologists' analysis. In fact, the international TIL working group recommended that instead of counting individual cells, TILs should be assessed as an area that the dense clusters cover in a whole section (Salgado et al., 2014). A benefit of the regional approach over cell-level analysis is fewer computational demands. No separate detection algorithm is required, and several cells can be analyzed simultaneously when a WSI is processed in patches.

Importantly, we also introduced an antibody-guided image annotation that can be broadly generalized into different supervised analysis settings. The annotation process requires two consecutively cut tissue sections and follows staining with H&E and a specific IHC marker, which stains the tissue entity of interest. To identify TILs, we used the pan-leukocyte marker CD45. IHC marker guided the annotation process of the H&E-stained section and aided detection and verification of small or otherwise unclear regions and thus

enabled more accurate and faster labeling. This provides an opportunity to automate the annotation process by image registration and IHC analysis. However, antibody-guided labeling can only be applied if a suitable marker exists. For instance, this approach cannot be used in case of necrosis but can be beneficial in annotation of entities such as mitoses, blood veins, and tumor tissue.

To summarize, in Study II we demonstrated the feasibility of TIL quantification in WSIs with regional tissue analysis. Our evaluation showed that the performance of the computerized assessment was comparable with pathologists' assessments. Additionally, we adapted the use of a pre-trained CNN in analysis of a digitized histological specimen and proposed a novel approach for improving image labeling via IHC.

Studies I and II focused on tumor tissue characterization via assessment of particular tumor features and the tumor microenvironment. Our results and work by others clearly show that computerized assessment of these characteristics is feasible and highly concordant with pathologist assessments. When incorporated in analysis of large patient cohorts, automated tumor tissue characterization, such as assessment of tumor viability and TILs, may provide a deeper understanding of tumor heterogeneity and the tumor microenvironment. Perhaps more importantly, machine-learning algorithms offer also novel opportunities to systematically analyze large patient data sets in a blinded fashion. These kinds of exploratory analyses may discover unobserved patterns and associations with outcomes of interest.

In Study III, we studied breast cancer patient outcome prediction with computer vision using a large nationwide patient cohort with TMA spots and clinical follow-up information. We hypothesized that using only the raw pixel values (the plain image) without any domain knowledge or introduction of fundamental tissue entities (i.e. cell types, tissue compartments, or distinction between cancerous and non-cancerous tissue) we could potentially identify prognostic signals that are complementary to traditional prognostic factors. Stratification of cancer patients into subgroups with different outcomes may guide treatment decision making and ultimately improve the understanding of cancer biology.

The limited availability of patient cohorts with tumor specimens and corresponding follow-up information is a challenge for the development of direct computerized outcome prediction. Nevertheless, patient outcome is largely influenced by tumor grade, and computerized grade classification has been studied broadly in different cancers, including brain (Ertosun & Rubin, 2015), kidney (Yeh et al., 2014), and prostate (D. Wang et al., 2015). Although cancer grade correlates with patient outcome, it is only an intermediate endpoint and might not completely capture the visual signs in the specimen that are relevant for the outcome. A seminal work in computerized outcome prediction (Beck et al., 2011) identified stromal features as independent prognostic factors for overall survival (HR=1.78) in breast cancer. Later, morphological features, prognostic for 8-year disease-free survival, were identified through a computerized analysis of H&E-stained specimens (J.-M. Chen et al., 2015). Another study combined image features with gene expression data for breast cancer patient prognostication (HR=1.70) (Popovici et al., 2016).

In Study III, we exploited a deep CNN in feature extraction and a feature-aggregation method that adjusts the local convolutional descriptors to the data domain. This yielded a risk stratification that turned out to be an independent predictor of breast cancer-specific survival (HR=2.06) when adjusted for relevant clinicopathological factors. These findings in breast cancer are supported by our work in colorectal cancer, where digital risk grouping predicted disease-specific survival (HR=1.89) when adjusted for a comprehensive set of covariables (Bychkov et al., 2018).

Subset analyses showed that DRS had differential survival profiles among patient subgroups. Notably, we found that patients with HER2-positive cancer were stratified with an over 40% difference in 10-year DSS, and that only grade-I cancers displayed a significant difference between the low and high DRS groups. In addition, survival profiles were significantly different in ER-, ER+, PR-, and HER2- and in both node-negative and node-positive cancers. This highlights the potential of data-driven approaches to discover novel patient groups for further analyses and for hypothesis generation.

A limitation of the suggested prediction method was the difficulty in fully explaining the source of the prediction. To compare the DRS with the current gold standard and to gain further insight of the DRS, we visually scored the same TMA spots that were studied. The visual scoring was performed by three experienced pathologists, whose assessments of the risk (low vs. high) were combined into a consensus score via majority voting. Correlation analysis revealed that the computerized prediction had a significant association with pleomorphism and tubule formation, whereas the visual risk score also captured mitoses, presence of necrosis, and TILs. This indicates that the DRS might partly capture the same tumor features that pathologists usually assess. Interestingly, survival analysis with the visual risk score and DRS as covariates implied both as independent predictive factors despite the overlap. In fact, the combined survival model resulted in a stronger discrimination than either of the risk scores alone, indicating that DRS holds complementary information to the pathologists' visual risk assessment.

Taken together, we demonstrated in Study III that direct outcome prediction with computer vision is not only feasible but can also provide independent predictive information complementary to clinicopathological factors and pathologists' visual risk scorings. Direct outcome prediction can thereby potentially reveal prognostically relevant visual features that may be undiscovered. However, studies on different cancer types and larger patient cohorts are still required for validation.

Computerized analysis of medical information has held the promise of improving patient care for decades (Schwartz, 1970). During the last few years, technological development in the field of machine learning has been faster than ever before. In particular, methods based on deep learning have repeatedly pushed the upper performance limit higher in many data domains and computational tasks. The two key components driving this development are open data sets and open computational frameworks. The frameworks are directly applicable for digital pathology by providing the complex algorithms in easy-to-use packages, which can be used to build computer-vision applications for analysis of histological specimens.

However, there are only a few high-quality data sets[5,6,7] available for the research community.

A superhuman performance level has been achieved in specific tasks using machine-learning algorithms. This means that computer software can solve a particular problem more precisely and faster than human experts. A famous example of such software is the algorithm developed by DeepMind that defeated the leading world champion in the game of Go (Silver et al., 2016, 2017), a decade earlier than anticipated. Rapid digitization of cancer specimens has enabled use of these same tools and algorithms to improve histological assessments. This may provide superhuman skills for future researchers and clinicians to better understand their data, ultimately leading to improved patient care.

[5] http://tupac.tue-image.nl/

[6] https://camelyon17.grand-challenge.org/

[7] https://www.kaggle.com/c/data-science-bowl-2018

7 CONCLUSIONS

In this doctoral thesis, we developed and evaluated computer-vision methods for characterization of tumor tissue and prediction of patient outcome. The main conclusions of our work are as follows:

1. Texture analysis can facilitate robust quantification of tumor necrosis.

2. Image descriptors extracted with a deep CNN can be adapted for characterization of breast cancer tissue into different compartments with accuracy comparable to pathologists' assessments.

3. IHC-guided image labeling facilitates accurate and fast ground-truth annotation.

4. Direct outcome prediction may identify novel prognostic image features that are complementary to traditional prognostic factors.

8 ACKNOWLEDGEMENTS

This study was performed at the Institute for Molecular Medicine Finland (FIMM), University of Helsinki during the years 2013-2018. I would like to acknowledge Professors Olli Kallioniemi and Jaakko Kaprio for acting as director of FIMM and for providing outstanding infrastructure during these years.

I am deeply grateful for the financial support I received from the Association for Pathology Informatics, Finnish Cancer Societies, Orion-Pharmos Research Foundation, Biomedicum Helsinki Foundation, Ida Montinin Foundation, Emil Aaltonen Foundation, and the Doctoral Programme in Biomedicine.

First, I would like to thank my supervisors Johan Lundin and Nina Linder for their guidance and advice during my PhD studies. It has been truly interesting to work with you, and I feel lucky that I have learned so much from you.

Jorma Laaksonen and Jorma Isola are acknowledged for acting as thesis committee members. I wish to thank you for encouragement, insightful discussion and feedback in our annual meetings. Also, the official thesis reviewers Johan Hartman and Claes Lundström are acknowledged for their expertise and effort to review this work. I wish to acknowledge Pekka Ruusuvuori for accepting to serve as the opponent of my defense, and Sampsa Hautaniemi for accepting the role of custos.

The work presented in this thesis is result of team work. Therefore, I would like to sincerely thank all the collaborators I have had pleasure to work with. Especially, I want to thank Clare Verrill, Heikki Joensuu, Kari Alitalo, Panu Kovanen, Tanja Holopainen, Stig Nordling and Teijo Pellinen for successful collaboration. Without your advice, knowledge, and expertise I would not have been able to carry out this work. Likewise, many others that I have had chance to work with in projects related to this thesis and outside it, are deeply acknowledged.

I have been fortunate to have great colleagues and friends around me during the years at FIMM. I want to acknowledge all the group members: Anne, Dima, Hakan, Klaus, Margarita, Micke, Oscar, and Yinhai for support and friendship. Thank you!

The entire community of FIMM is acknowledged for the unique working environment. It has been truly fascinating to work in such an inspiring community, packed with talented and ambitious people. Particularly, I want to thank Andrew, Dimi, Heikki, Jarno, Lassi, Oscar, Pyry, Sami, Teijo, and Vesa for vital peer-support, shared sense of humor, and friendship – without what this thesis would never have been finalized. Importantly, the FIMM Unscientific Coffee Klubben and HaSB are acknowledged for wholesome activities that helped me to maintain both mental and physical fitness. Furthermore, I wish to thank Riku and Risto for the countless hours spent at climbing gyms and trails of central park, giving me much needed balance between work and play.

Finally, I would like to thank my family. Thank you mum and dad, Irma and Ilkka, for all the encouragement. I also wish to thank my sister Riikka and brother Jaakko for their support and inspiration.

Most of all, I want to thank Elina for patience, encouragement, support, and energy you gave me throughout this PhD process. Thank you!

Helsinki, 2018

Riku Turkki

9 REFERENCES

Achanta R., Shaji A., Smith K., Lucchi A., Fua P., & Süsstrunk S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(11).

Al-Janabi S., Huisman A., Nap M., Clarijs R., & van Diest P. J. (2012). Whole slide images as a platform for initial diagnostics in histopathology in a medium-sized routine laboratory. *Journal of Clinical Pathology*, *65*(12).

Al-Kofahi Y., Lassoued W., Lee W., & Roysam B. (2010). Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Bio-Medical Engineering*, *57*(4).

Al Habeeb A., Evans A., & Ghazarian D. (2012). Virtual microscopy using whole-slide imaging as an enabler for teledermatopathology: A paired consultant validation study. *Journal of Pathology Informatics*, *3*.

Awan R., Sirinukunwattana K., Epstein D., Jefferyes S., Qidwai U., Aftab Z., Mujeeb I., Snead D., & Rajpoot N. (2017). Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images. *Scientific Reports*, *7*(1).

Basavanhally A., Ganesan S., Feldman M., Shih N., Mies C., Tomaszewski J., & Madabhushi A. (2013). Multi-field-of-view framework for distinguishing tumor grade in ER+ breast cancer from entire histopathology slides. *IEEE Transactions on Bio-Medical Engineering*, *60*(8).

Bauer T. W., & Slaw R. J. (2014). Validating Whole-Slide Imaging for Consultation Diagnoses in Surgical Pathology. *Archives of Pathology & Laboratory Medicine*, *138*(11).

Beck A. H., Sangoi A. R., Leung S., Marinelli R. J., Nielsen T. O., van de Vijver M. J., West R. B., van de Rijn M., & Koller D. (2011). Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival. *Science Translational Medicine*, *3*(108).

Bejnordi B. E., Veta M., Johannes van Diest P., van Ginneken B., Karssemeijer N., Litjens G., van der Laak J. A. W. M., Hermsen M., Manson Q. F., … Venâncio R. (2017). Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*, *318*(22).

Bhargava R., & Madabhushi A. (2016). Emerging Themes in Image Informatics and Molecular Analysis for Digital Pathology. *Annual Review of Biomedical Engineering*, *18*.

Bychkov D., Linder N., Turkki R., Nordling S., Kovanen P. E., Verrill C., Walliander M., Lundin M., Haglund C., & Lundin J. (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific Reports*, *8*(1).

Carey L. A., Perou C. M., Livasy C. A., Dressler L. G., Cowan D., Conway K., Karaca G., Troester M. A., Chiu K. T., Edmiston S., Deming S. L., Geradts J., Cheang M. C. U., Nielsen T. O., Moorman P. G., Earp H. S., & Millikan R. C. (2006). Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *Journal of the American Medical Association*.

Caruso R., Parisi A., Bonanno A., Paparo D., Quattrocchi E., Branca G., Scardigno M.,

& Fedele F. (2011). Histologic coagulative tumour necrosis as a prognostic indicator of aggressiveness in renal, lung, thyroid and colorectal carcinomas: A brief review. *Oncology Letters*, *3*(1).

Chan J. K. C. (2014). The Wonderful Colors of the Hematoxylin–Eosin Stain in Diagnostic Surgical Pathology. *International Journal of Surgical Pathology*, *22*(1).

Chatfield K., Simonyan K., Vedaldi A., & Zisserman A. (2014). Return of the Devil in the Details: Delving Deep into Convolutional Nets.

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

Chen J.-M., Li Y., Xu J., Gong L., Wang L.-W., Liu W.-L., & Liu J. (2017). Computer-aided prognosis on breast cancer with hematoxylin and eosin histopathology images: A review. *Tumor Biology*, *39*(3).

Chen J.-M., Qu A.-P., Wang L.-W., Yuan J.-P., Yang F., Xiang Q.-M., Maskey N., Yang G.-F., Liu J., & Li Y. (2015). New breast cancer prognostic factors identified by computer-aided image analysis of HE stained histopathology images. *Scientific Reports*, *5*.

Cimpoi M., Maji S., Kokkinos I., & Vedaldi A. (2016). Deep Filter Banks for Texture Recognition, Description, and Segmentation. *International Journal of Computer Vision*.

Cireşan D. C., Giusti A., Gambardella L. M., & Schmidhuber J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. *Medical Image Computing and Computer-Assisted Intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, *16*(Pt 2).

Coons A. H., Creech H. J., & Jones R. N. (1941). Immunological Properties of an Antibody Containing a Fluorescent Group. *Experimental Biology and Medicine*.

Cortes C., & Vapnik V. (1995). Support-vector networks. *Machine Learning*, *20*(3).

Cox D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*.

Djuric U., Zadeh G., Aldape K., & Diamandis P. (2017). Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care. *Npj Precision Oncology*.

Elston C. W., & Ellis I. O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, *19*(5).

Epstein J. I., Egevad L., Amin M. B., Delahunt B., Srigley J. R., & Humphrey P. A. (2016). The 2014 international society of urological pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma definition of grading patterns and proposal for a new grading system. *American Journal of Surgical Pathology*.

Ertosun M. G., & Rubin D. L. (2015). Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of

convolutional neural networks. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2015*.

Esteva A., Kuprel B., Novoa R. A., Ko J., Swetter S. M., Blau H. M., & Thrun S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639).

Fan R.-E., Chang K.-W., Hsieh C.-J., Wang X.-R., & Lin C.-J. (2008). LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, *9*.

Farahani N., & Pantanowitz L. (2015). Overview of Telepathology. *Surgical Pathology Clinics*, *8*(2).

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

Fletcher C. D. M. (2013). *Diagnostic histopathology of tumors*.

Fogel I., & Sagi D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*.

Fouad S., Randell D., Galton A., Mehanna H., & Landini G. (2017). Unsupervised morphological segmentation of tissue compartments in histopathological images. *PLOS ONE*, *12*(11).

Fridman W. H., Zitvogel L., Sautès–Fridman C., & Kroemer G. (2017). The immune contexture in cancer prognosis and treatment. *Nature Reviews Clinical Oncology*, *14*(12).

Fritz P., Klenk S., Goletz S., Gerteis A., Simon W., Brinkmann F., Heidemann E., Lütttgen E., Ott G., Alscher M. D., Schwab M., & Dippon J. (2010). Clinical impacts of histological subtyping primary breast cancer. *Anticancer Research*.

Gönen M., & Heller G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, *92*(4).

Green D. R. (2011). *Means to an end : apoptosis and other cell death mechanisms*.

Griffin J., & Treanor D. (2017). Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology*, *70*(1).

Gulshan V., Peng L., Coram M., Stumpe M. C., Wu D., Narayanaswamy A., Venugopalan S., Widner K., Madams T., Cuadros J., Kim R., Raman R., Nelson P. C., Mega J. L., & Webster D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, *316*(22).

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

Hamilton P. W., Bankhead P., Wang Y., Hutchinson R., Kieran D., McArt D. G., James J., & Salto-Tellez M. (2014). Digital pathology and image analysis in tissue biomarker research. *Methods*, *70*(1).

Hanahan D., & Weinberg R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5).

Haralick R. M., Dinstein I., & Shanmugam K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*.

Hockel M., & Vaupel P. (2001). Tumor Hypoxia: Definitions and Current Clinical, Biologic, and Molecular Aspects. *JNCI Journal of the National Cancer Institute*, *93*(4).

Hoff E. R., Tubbs R. R., Myles J. L., & Procop G. W. (2002). HER2/neu amplification in breast cancer: Stratification by tumor type and grade. *American Journal of Clinical Pathology*.

Homeyer A., Schenk A., Arlt J., Dahmen U., Dirsch O., & Hahn H. K. (2013). Practical quantification of necrosis in histological whole-slide images. *Computerized Medical Imaging and Graphics*.

Jafari-Khouzani K., & Soltanian-Zadeh H. (2003). Multiwavelet grading of pathological images of prostate. *IEEE Transactions on Biomedical Engineering*, *50*(6).

Janowczyk A., & Madabhushi A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, *7*(1).

Jia Deng, Wei Dong, Socher R., Li-Jia Li, Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255).

Junqueira L., & Carneiro J. (2005). Basic Histology: Text & Atlas. *Statrefcom*.

Kaplan E. L., & Meier P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, *53*(282).

Kather J. N., Weis C.-A., Bianconi F., Melchers S. M., Schad L. R., Gaiser T., Marx A., & Zöllner F. G. (2016). Multi-class texture analysis in colorectal cancer histology. *Scientific Reports*.

Khan A. M., Rajpoot N., Treanor D., & Magee D. (2014). A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution. *IEEE Transactions on Biomedical Engineering*, *61*(6).

Klette R. (2014). *Concise Computer Vision*.

Kononen J., Bubendorf L., Kallioniemi A., Bärlund M., Schraml P., Leighton S., Torhorst J., Mihatsch M. J., Sauter G., & Kallioniemi O. P. (1998). Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Medicine*, *4*(7).

Le Q. V., Le Q. V., Han J., Spellman P., Borowsky E., & Parvin B. (2012). Learning invariant features of tumor signatures. *IN ISBI*.

LeCun Y., Bengio Y., & Hinton G. (2015). Deep learning. *Nature*, *521*(7553).

Lee S., & Margolin K. (2012). Tumor-infiltrating lymphocytes in melanoma. *Current Oncology Reports*, *14*(5).

Letai A. (2017). Functional precision cancer medicine—moving beyond pure genomics.

Nature Medicine, *23*(9).

Li C. I., Daling J. R., & Malone K. E. (2003). Incidence of invasive breast cancer by hormone receptor status from 1992 to 1998. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*.

Linder N., Konsti J., Turkki R., Rahtu E., Lundin M., Nordling S., Haglund C., Ahonen T., Pietikäinen M., & Lundin J. (2012). Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagnostic Pathology*, *7*(1).

Litjens G., Kooi T., Bejnordi B. E., Setio A. A. A., Ciompi F., Ghafoorian M., van der Laak J. A. W. M., van Ginneken B., & Sánchez C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*.

Loi S., Michiels S., Salgado R., Sirtaine N., Jose V., Fumagalli D., Kellokumpu-Lehtinen P.-L., Bono P., Kataja V., Desmedt C., Piccart M. J., Loibl S., Denkert C., Smyth M. J., Joensuu H., & Sotiriou C. (2014). Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Annals of Oncology*, *25*(8).

Luen S. J., Savas P., Fox S. B., Salgado R., & Loi S. (2017). Tumour-infiltrating lymphocytes and the emerging role of immunotherapy in breast cancer. *Pathology*.

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

Matos L. L. de, Trufelli D. C., de Matos M. G. L., & da Silva Pinhal M. A. (2010). Immunohistochemistry as an important tool in biomarkers detection and clinical practice. *Biomarker Insights*, *5*.

Meyer J. S., Alvarez C., Milikowski C., Olson N., Russo I., Russo J., Glass A., Zehnbauer B. A., Lister K., & Parwaresch R. (2005). Breast carcinoma malignancy grading by Bloom–Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. *Modern Pathology*, *18*(8).

Mlecnik B., Bindea G., Kirilovsky A., Angell H. K., Obenauf A. C., Tosolini M., Church S. E., Maby P., Vasaturo A., Angelova M., Fredriksen T., Mauger S., Waldner M., Berger A., Speicher M. R., Pagès F., Valge-Archer V., & Galon J. (2016). The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis. *Science Translational Medicine*, *8*(327).

Muley T. R., Herth F. J., Schnabel P. A., Dienemann H., & Meister M. (2012). From tissue to molecular phenotyping: pre-analytical requirements heidelberg experience. *Translational Lung Cancer Research*, *1*(2).

Nawaz S., Heindl A., Koelble K., & Yuan Y. (2015). Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Modern Pathology*, *28*(6).

Nayak N., Chang H., Borowsky A., Spellman P., & Parvin B. (2013). Classification of tumor histopathology via sparse feature learning. In *2013 IEEE 10th International Symposium on Biomedical Imaging* (Vol. 2013, pp. 410–413).

Nicolini A., Ferrari P., & Duffy M. J. (2017). Prognostic and predictive biomarkers in

breast cancer: Past, present and future. *Seminars in Cancer Biology*.

Ojala T., Pietikäinen M., & Mäenpää T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, *24*(7).

Oldenhuis C. N. A. M., Oosting S. F., Gietema J. A., & de Vries E. G. E. (2008). Prognostic versus predictive value of biomarkers in oncology. *European Journal of Cancer*, *44*(7).

Panagiotakis C., Ramasso E., & Tziritas G. (2010). Lymphocyte segmentation using the transferable belief model.

Pantanowitz L., Valenstein P. N., Evans A. J., Kaplan K. J., Pfeifer J. D., Wilbur D. C., Collins L. C., & Colgan T. J. (2011). Review of the current state of whole slide imaging in pathology. *Journal of Pathology Informatics*, *2*.

Perronnin F., & Dance C. (2007). Fisher Kernels on Visual Vocabularies for Image Categorization. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8).

Perronnin F., Sánchez J., & Mensink T. (2010). Improving the Fisher Kernel for Large-Scale Image Classification (pp. 143–156).

Pietikäinen M., Hadid A., Zhao G., & Ahonen T. (2011). Computer Vision Using Local Binary Patterns. In *Computer Vision Using Local Binary Patterns*.

Pollheimer M. J., Kornprat P., Lindtner R. A., Harbaum L., Schlemmer A., Rehak P., & Langner C. (2010). Tumor necrosis is a new promising prognostic factor in colorectal cancer. *Human Pathology*, *41*(12).

Popovici V., Budinská E., Čápková L., Schwarz D., Dušek L., Feit J., & Jaggi R. (2016). Joint analysis of histopathology image features and gene expression in breast cancer. *BMC Bioinformatics*, *17*(1).

Rakha E. A., El-Sayed M. E., Lee A. H. S., Elston C. W., Grainge M. J., Hodi Z., Blamey R. W., & Ellis I. O. (2008). Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*.

Robbins S. L., Kumar V., & Cotran R. S. (2010). *Robbins and Cotran pathologic basis of disease.*

Robertson S., Azizpour H., Smith K., & Hartman J. (2017). Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. *Translational Research : The Journal of Laboratory and Clinical Medicine*.

Rocha R., Vassallo J., Soares F., Miller K., & Gobbi H. (2009). Digital slides: Present status of a tool for consultation, teaching, and quality control in pathology. *Pathology - Research and Practice*, *205*(11).

Ruffell B., Au A., Rugo H. S., Esserman L. J., Hwang E. S., & Coussens L. M. (2012). Leukocyte composition of human breast cancer. *Proceedings of the National Academy of Sciences*.

Salgado R., Denkert C., Demaria S., Sirtaine N., Klauschen F., Pruneri G., Wienert S., Van den Eynden G., Baehner F. L., Penault-Llorca F., Perez E. A., Thompson E. A., Symmans W. F., Richardson A. L., Brock J., Criscitiello C., Bailey H.,

Ignatiadis M., Floris G., Sparano J., Kos Z., Nielsen T., Rimm D. L., Allison K. H., Reis-Filho J. S., Loibl S., Sotiriou C., Viale G., Badve S., Adams S., Willard-Gallo K., & Loi S. (2014). The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO*, *26*(2).

Santoiemma P. P., & Powell D. J. (2015). Tumor infiltrating lymphocytes in ovarian cancer. *Cancer Biology & Therapy*, *16*(6).

Savas P., Salgado R., Denkert C., Sotiriou C., Darcy P. K., Smyth M. J., & Loi S. (2015). Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nature Reviews Clinical Oncology*.

Schmidhuber J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*.

Schwartz W. B. (1970). Medicine and the Computer: The Promise and Problems of Change (pp. 321–335).

Sharma H., Zerbe N., Klempert I., & Hellwich O. (2017). Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Computerized Medical Imaging and Graphics*, *61*.

Sharma H., Zerbe N., Klempert I., Lohmann S., Lindequist B., Hellwich O., & Hufnagl P. (2015). Appearance-based necrosis detection using textural features and SVM with discriminative thresholding in histopathological whole slide images. In *2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)* (pp. 1–6).

Sheikhzadeh F., Ward R. K., van Niekerk D., & Guillaud M. (2018). Automatic labeling of molecular biomarkers of immunohistochemistry images using fully convolutional networks. *PLOS ONE*, *13*(1).

Signolle N., Plancoulaine B., Herlin P., & Revenu M. (2008). Texture-based multiscale segmentation: Application to stromal compartment characterization on ovarian carcinoma virtual slides. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Signolle N., Revenu M., Plancoulaine B., & Herlin P. (2010). Wavelet-based multiscale texture segmentation: Application to stromal compartment characterization on virtual slides. *Signal Processing*.

Silver D., Huang A., Maddison C. J., Guez A., Sifre L., van den Driessche G., Schrittwieser J., Antonoglou I., Panneershelvam V., Lanctot M., Dieleman S., Grewe D., Nham J., Kalchbrenner N., Sutskever I., Lillicrap T., Leach M., Kavukcuoglu K., Graepel T., & Hassabis D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587).

Silver D., Schrittwieser J., Simonyan K., Antonoglou I., Huang A., Guez A., Hubert T., Baker L., Lai M., Bolton A., Chen Y., Lillicrap T., Hui F., Sifre L., van den Driessche G., Graepel T., & Hassabis D. (2017). Mastering the game of Go without human knowledge. *Nature*, *550*(7676).

Simonyan K., & Zisserman A. (2014). Very Deep Convolutional Networks for Large-

Scale Image Recognition.

Sirinukunwattana K., Pluim J. P. W., Chen H., Qi X., Heng P.-A., Guo Y. B., Wang L. Y., Matuszewski B. J., Bruni E., Sanchez U., Böhm A., Ronneberger O., Cheikh B. Ben, Racoceanu D., Kainz P., Pfeiffer M., Urschler M., Snead D. R. J., & Rajpoot N. M. (2017). Gland segmentation in colon histology images: The glas challenge contest. *Medical Image Analysis*, *35*.

Sistrunk W. E., & Maccarty W. C. (1922). Life expectancy following radical amputation for carcinoma of the breast: a clinical and pathologic study of 218 cases. *Annals of Surgery*, *75*(1).

Snead D. R. J., Tsang Y.-W., Meskiri A., Kimani P. K., Crossman R., Rajpoot N. M., Blessing E., Chen K., Gopalakrishnan K., Matthews P., Momtahan N., Read-Jones S., Sah S., Simmons E., Sinha B., Suortamo S., Yeo Y., El Daly H., & Cree I. A. (2016). Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology*, *68*(7).

Stathonikos N., Veta M., Huisman A., & van Diest P. J. (2013). Going fully digital: Perspective of a Dutch academic pathology lab. *Journal of Pathology Informatics*, *4*(1).

Strimbu K., & Tavel J. A. (2010). What are biomarkers? *Current Opinion in HIV and AIDS*, *5*(6).

Sun Z., Aubry M.-C., Deschamps C., Marks R. S., Okuno S. H., Williams B. A., Sugimura H., Pankratz V. S., & Yang P. (2006). Histologic grade is an independent prognostic factor for survival in non–small cell lung cancer: An analysis of 5018 hospital- and 712 population-based cases. *The Journal of Thoracic and Cardiovascular Surgery*, *131*(5).

Swinson D. E. B., Jones J. L., Richardson D., Cox G., Edwards J. G., & O'Byrne K. J. (2002). Tumour necrosis is an independent prognostic marker in non-small cell lung cancer: correlation with biological variables. *Lung Cancer (Amsterdam, Netherlands)*, *37*(3).

Tabesh A., Teverovskiy M., Pang H.-Y., Kumar V. P., Verbel D., Kotsianti A., & Saidi O. (2007). Multifeature prostate cancer diagnosis and Gleason grading of histological images. *IEEE Transactions on Medical Imaging*, *26*(10).

Tavassoéli F., & Devilee P. (eds. (2003). *Pathology and Genetics of Tumours of the Breast and Female Genital Organs*.

Theodoridis S., & Koutroumbas K. (2009). *Pattern recognition*.

Thorstenson S., Molin J., & Lundström C. (2014). Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006-2013. *Journal of Pathology Informatics*, *5*(1).

Toikkanen S., & Joensuu H. (1990). Prognostic factors and long-term survival in breast cancer in a defined urban population. *APMIS*.

Tredan O., Galmarini C. M., Patel K., & Tannock I. F. (2007). Drug Resistance and the Solid Tumor Microenvironment. *JNCI Journal of the National Cancer Institute*, *99*(19).

Tuominen V. J., Ruotoistenmäki S., Viitanen A., Jumppanen M., & Isola J. (2010).

ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Research : BCR*, *12*(4).

Turkki R., Linder N., Holopainen T., Wang Y., Grote A., Lundin M., Alitalo K., & Lundin J. (2015). Assessment of tumour viability in human lung cancer xenografts with texture-based image analysis. *Journal of Clinical Pathology*, *68*(8).

Turkki R., Linder N., Kovanen P. E., Pellinen T., & Lundin J. (2016a). Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. *Journal of Pathology Informatics*, *7*(1).

Turkki R., Linder N., Kovanen P. E., Pellinen T., & Lundin J. (2016b). Identification of immune cell infiltration in hematoxylin-eosin stained breast cancer samples: texture-based classification of tissue morphologies. *SPIE Medical Imaging*.

Vaynrub M., Taheri N., Ahlmann E. R., Yao C., Fedenko A. N., Allison D. C., Chawla S. P., & Menendez L. R. (2015). Prognostic value of necrosis after neoadjuvant therapy for soft tissue sarcoma. *Journal of Surgical Oncology*.

Vedaldi A., & Fulkerson B. (2010). VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the international conference on Multimedia* (pp. 1469–1472).

Vedaldi A., & Lenc K. (2014). MatConvNet - Convolutional Neural Networks for MATLAB.

Vedaldi A., & Zisserman A. (2012). Sparse kernel approximations for efficient classification and detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 2320–2327).

Vestjens J. H. M. J., Pepels M. J., De Boer M., Borm G. F., Van Deurzen C. H. M., Van Diest P. J., Van Dijck J. A. A. M., Adang E. M. M., Nortier J. W. R., Rutgers E. J. T., Seynaeve C., Menke-Pluymers M. B. E., Bult P., & Tjan-Heijnen V. C. G. (2012). Relevant impact of central pathology review on nodal classification in individual breast cancer patients. *Annals of Oncology*.

Veta M., Pluim J. P. W., van Diest P. J., & Viergever M. A. (2014). Breast Cancer Histopathology Image Analysis: A Review. *IEEE Transactions on Biomedical Engineering*, *61*(5).

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., & Pluim J. P. W. (2015). Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis*, *20*(1).

Vodovnik A. (2016). Diagnostic time in digital pathology: A comparative study on 400 cases. *Journal of Pathology Informatics*, *7*(1).

Wang D., Foran D. J., Ren J., Zhong H., Kim I. Y., & Qi X. (2015). Exploring automatic prostate histopathology image Gleason grading via local structure modeling. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and*

Biology Society. Annual Conference, *2015*.

Wang H., Cruz-Roa A., Basavanhally A., Gilmore H., Shih N., Feldman M., Tomaszewski J., Gonzalez F., & Madabhushi A. (2014). Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging (Bellingham, Wash.)*, *1*(3).

Weinberg R. A. (2007). *The biology of cancer*.

Whiteside T. L. (2008). The tumor microenvironment and its role in promoting tumor growth. *Oncogene*, *27*(45).

Xu J., Luo X., Wang G., Gilmore H., & Madabhushi A. (2016). A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing*, *191*.

Yeh F.-C., Parwani A. V, Pantanowitz L., & Ho C. (2014). Automated grading of renal cell carcinoma using whole slide imaging. *Journal of Pathology Informatics*, *5*(1).

Yu K.-H., Zhang C., Berry G. J., Altman R. B., Ré C., Rubin D. L., & Snyder M. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature Communications*, *7*.

10 ORIGINAL PUBLICATIONS