Identificación automática de marcadores patológicos en imágenes de histopatología

David Edmundo Romo Bucheli

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería Eléctrica y Electrónica
Bogotá D.C., Colombia
2017

Automatic identification of pathological markers in histopathological images

David Edmundo Romo Bucheli

In fulfillment of the requirements for the degree of:
Doctor en Ingeniería - Ingeniería Eléctrica

Advisor:
Eduardo Romero Castro, Ph.D.

Research Field:
Applied Computing - Image and Video Processing
Research Group:
Computer Imaging and Medical Applications Laboratory - CIM@LAB

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería Eléctrica y Electrónica
Bogotá D.C., Colombia
2017

A mis padres, Luis Edmundo y Esperanza

Acknowledgements

First of all, I want to express my gratitude to my advisor, Professor Eduardo Romero for his patience, guidance and wise advice during all my doctoral studies. I also would like to thank specially to the Professor Anant Madabhushi for his advice and important contributions in my work and the opportunity to carry an internship at the CCIPD laboratory in Case Western Reserve University.

Thanks to all people in Cim@lab research group that have contributed at different scales in my work and personal enrichment. To my great friend Fabio Martinez and also to my old masters degree teammates for the broad support received during this process. Thanks also to all my co-authors: Andrew Janowczyk, Angel Cruz, Germán Corredor, Dr. Hannah Gilmore, Jonathan Tarquino, Juan David García, Pablo Arbeláez, Raúl Celis, Ricardo Moncayo and Dr. Viviana Arias for their unvaluable help and feedback.

I thank to the *Universidad Nacional de Colombia* that have sponsored my Ph.D. formation with the scholarship: *Becas para estudiantes sobresalientes de posgrado.*

Abstract

The inter and intra subject variability is a common problem in several tasks associated to the examination of histopathological samples. This variability might hinder the evaluation of cancerous diseases. The development of automatic image analysis techniques and computerized aided diagnostic tools in pathology aims to reduce the impact of such variability by offering quantitative measurements and estimations. These measurements allow an accurate evaluation and classification of the diseases in virtual slide images. The main problem addressed in this thesis is evaluating the correlation of the automated identification of pathological markers with cancer malignancy and aggresivenes. Hence, a set of classifier models are trained to detect known pathological patterns. The classifiers are then used to quantify the presence of the pathological markers. Finally, the resulting measurements are correlated with the cancer risk recurrence. Results show that the automated detectors are able to quantify patterns that show differences across several cancer risk groups.

Keywords: Histopathology, Digital pathology, Pathological marker

Resumen

La variabilidad inter e intra sujeto es un problema frecuente en muchas tareas asociadas al exámen de muestras histopatológicas. Esta variabilidad puede incidir negativamente en la evaluación de patologías relacionadas con el cáncer. El desarrollo de técnicas para el análisis automático de imágenes y de herramientas de soporte al diagnóstico en patología tiene como objetivo reducir el impacto de la variabilidad inter/intra sujeto mediante la obtención de medidas y estimaciones cuantitativas. Estas medidas permiten una evaluación y clasificación más precisa de las enfermedades observables en láminas virtuales. El principal problema abordado en esta tesis consiste en evaluar la correlación de la identificación automática de marcadores patológicos con la agresividad del cancer. Así, un conjunto de clasificadores son entrenados para detectar marcadores patológicos conocidos. Los clasificadores son posteriormente usados para cuantificar la presencia de los marcadores patológicos. Finalmente, las mediciones resultantes son correlacionadas con el riesgo de recurrencia del cáncer. Los resultados muestran que los detectores automáticos son capaces de cuantificar los patrones que muestran diferencias entre diferentes grupos de riesgo.

Palabras clave: Histopatología, Patología digital, Marcador patológico

Contents

Acknowledgements iv

Abstract v

1. Introduction 2

2. Automated Tubule Nuclei Quantification and Correlation with Oncotype DX risk categories in ER+ Breast Cancer Whole Slide Images 8

2.1. Introduction . . . 8
2.2. Methodology . . . 11
2.2.1. Nuclei detection . . . 11
2.2.2. Curating the Learning Set . . . 11
2.2.3. Independent testing of the DNN classifier . . . 13
2.3. Experimental Design . . . 14
2.3.1. Data Description . . . 14
2.3.2. Correlation with ODX risk groups via t-test analyses . . . 14
2.3.3. Correlation with ODX risk groups via ROC analysis . . . 15
2.4. Results . . . 16
2.4.1. Correlation with ODX and BR risk categories via t-test analyses . . . 16
2.4.2. Correlation with ODX and BR risk categories via ROC curve . . . 19
2.5. Concluding Remarks . . . 20

3. A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in Estrogen Receptor Positive Breast Cancers 21

3.1. Background . . . 22
3.2. Experimental Methods . . . 23
3.2.1. DNN for mitotic identification . . . 24
3.2.2. Support Vector Machine for BCa risk stratification . . . 25
3.3. Experimental Design . . . 25
3.3.1. Whole Slide Data Description . . . 25
3.3.2. Experiments and statistical analysis . . . 26

3.4. Experimental Results and Discussion 28
3.4.1. Statistical Analysis 28
3.4.2. Support Vector Machine Classification Results 30
3.5. Concluding Remarks 30

4. Conclusions **32**

Bibliography **34**

A. Appendix: Virtual Slide Mosaicing Using Feature Descriptors and a Registration Consistency Measure **42**

B. Appendix: A Discriminant Multi-Scale Histopathology Descriptor using Dictionary Learning **51**

C. Appendix: Blind colour separation of H&E stained histological images by linearly transforming the colour space **58**

D. Appendix: A Grading Strategy for Nuclear Pleomorphism in Histopathological Breast Cancer Images Using a Bag of Features (BOF) **71**

E. Appendix: Identifying Histological Concepts on Basal Cell Carcinoma Images using Nuclei based Sampling and Multi-scale descriptors **80**

F. Appendix: Nuclei Graph Local Features for Basal Cell Carcinoma Classication in Whole Slide Images **85**

G. Appendix: Supplementary File - Automated Tubule Nuclei Quantification and Correlation with Oncotype DX risk categories in ER+ Breast Cancer Whole Slide Images **93**

H. Appendix: Supplemenatary File A - Classifying Breast Cancer cases using only mean mitosis count per 10 High Power Fields (HPFs) **101**

I. Appendix: Supplemenatary File B - Supplementary File B - Mitosis detector performance in a set of high power fields **104**

1. Introduction

A virtual microscope system is comprised by a set of computational tools that allows the exploration and analysis of digitized slides (Virtual slides). The virtual slides contain a large amount of information that require an efficient interface and adequate storage. Deployment of virtual microscopy (VM) systems in the last decade has had a wide impact in medical education and training [2, 17, 25, 28, 30, 32, 36]. Unlike the pervasive support of technologies in these specific medical applications, integration of computerized tools into the routine diagnostic pathological workflow has been limited, likely because these systems require specialized facilities related with acquisition, digitization, storage and visualization of the histological samples. Moreover, strategies and tools providing diagnostic assistance do not offer criteria that aids the interpretative task in the analysis of histophatological images.

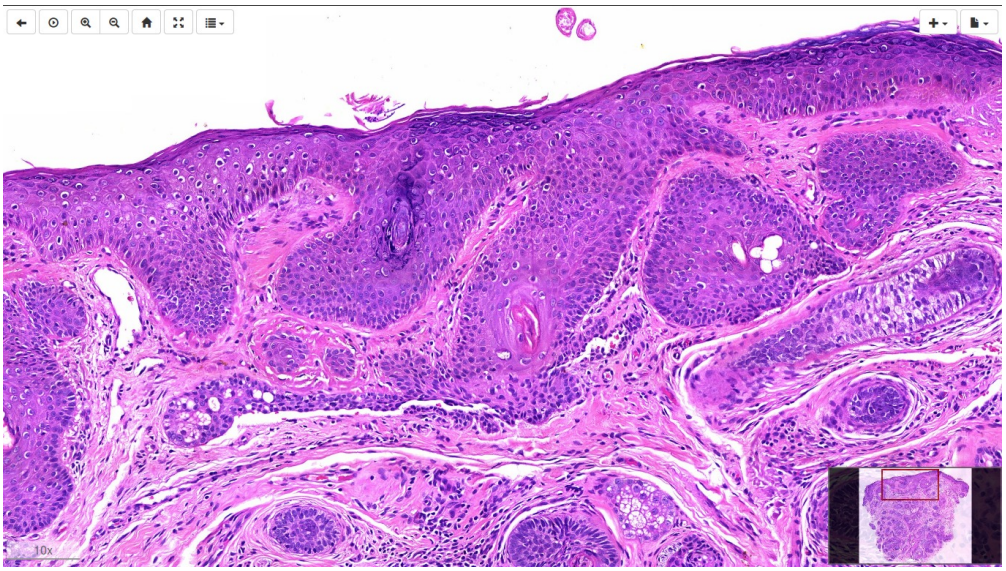

Figure 1-1.: Virtual microscopy system graphics interface. Navigation tools allow panning and zooming operations onto the virtual slides mimicking the operations available in a light microscope. Additional features, such as sharing and annotation tools, are highly regarded. However, the integration of computerized aiding tools to help the pathologist analysis is still lacking.

Inter and intra subject variability is a crucial problem in several diagnostic tasks. This variability inevitably introduces noise and biases evaluation of difficult pathological entities.

The development of automatic analysis and computerized aided diagnostic tools aims to reduce the impact of such variability by offering quantitative measurements and estimations.

Research Problem

From a visual examination standpoint, many pathological entities are hardly distinguishable. This difficulty is even worst when establishing the malignancy grade of cancer diseases, case in which architectural and cellular distortions are present. Identification of visual features characterizing such changes (pathological markers) are used as evidence of the medical diagnosis, yet the visual signs can be very subtle and definitely dependent on the pathologist's experience.

Research Question

The research question herein addressed was: *How to identify, in an efficient and reproducible way, visual patterns that allow the discrimination of diseases with different risk/aggressiveness?*

Virtual microscopy and variability in histopathology analysis

Histopathology is the study of pathological signs using the microscopic examination of a biopsy or histological specimen that has been previously processed and spread onto a glass slide [4]. Visualization of cellular structures by microscopical inspection is achieved with histological dyes that can bind to several microscopical structures in the tissue. The Hematoxylin and Eosin (the H&E stain) are the dyes most frequently used for pathological examination. The hematoxylin has more affinity for chemical components present in the cellular nucleus, staining the nuclei with a dark blue color, while the eosin is commonly bounded to chemicals present in the cytoplasm and connective tissue. These elements are commonly observed in a pink/red color. Work6Hours In spite of the avalanche of medical imaging modalities (such as magnetic resonance imaging (MRI) or computerized tomography (CT), among others) providing non-invasive examination of the pathological lesions, histological examination remains the gold standard for diagnosis of a considerable number of diseases, including the entire set of cancer diseases [4].

An histopathological examination is performed by an expert pathologist. This analysis consists in a subjective evaluation driven by years of medical training [12, 18, 42]. In general, a pathologist inspects the tissue structure, the cell distribution, and the cell shape and size of the pathological sample [54]. Afterwards, the expert reaches a decision about the tissue anomalies associated to benign or malignant changes. This process is laborious and, as said before, hampered by inter/intra subject variability. The computerized tools aim to reduce the impact of this variability by providing the quantitative analysis of several tissular features [33].

Automatic analysis of histopathological images

Design and implementation of systems that support pathological diagnosis, require the use of digital image processing techniques such as: image segmentation, feature extraction algorithms or classification algorithms among others [9]

The visual feature extraction at cellular level allows a nuclei/cell quantification of morphological characteristics. These quantitative measurements can determine anomalies and their malignancy grade by quantifying local features, but with no special consideration about the spatial distribution of the cells. For each nucleus/cell, visual features associated to color, orientation, intensity and shape (nuclei area, perimeter, eccentricity, among others) can be extracted [22, 11, 51, 24, 3]. Likewise, at tissular level, features correlating with spatial cell distribution such as texture information or fractal descriptors capture and describe some pathological entities [40, 23].

Some applications of the previously mentioned algorithms, integrated to systems supporting diagnostic decisions have also been proposed. Some pathologies include the neuroblastoma detection [34, 63], the assisted analysis of renal cancer [68], prostate cancer detection [24] and breast cancer [51].

Contributions and academic products

This thesis presents some contributions to the virtual microscopy and digital pathology fields. A main contribution corresponds to the study of the prognostic use of automated pathological markers in cancer disease. Machine learning algorithms (Deep neural networks and bag of features representations) were used to assess key histological criteria in digitized histological samples. Afterwards, quantification of these histological features was compared with either manual visual inspection grading or alternative gene test scores (which are directly correlated with cancer recurrence risk). Additional contributions are a novel method that normalizes the color stains in microscopical studies, microscopical image stitching, and the construction of graph local features that describe complex histopathological patterns.

The academic products of this work include three (3) accepted journal papers and six (6) conference papers:

Pre-processing in virtual microscopy

Visualising and analysing histological images require several pre-processing operations. This issues was studied by a set of publications in which we developed a colour separation of the hematoxylin and eosin channels from the H& E stain [13]. Since the image depends on the concentration of these chemicals, the fundamental application of this technique is the nuclear detection/segmentation and color normalization of digital slides. Also, we proposed

a robust stitching method for construction of whole slide images acquired by a motorized microscope[56].

- **David E. Romo**, Jonathan Tarquino, Juan D. García-Arteaga and Eduardo Romero. *Virtual slide mosaicing using feature descriptors and a registration consistency measure*, Proceedings of the IX International Seminar on Medical Information Processing and Analysis, 2013.

- Raul Celis, **David Romo** and Eduardo Romero. *Blind colour separation of H&E stained histological images by linearly transforming the colour space*, Journal of Microscopy, vol. 260, no 3, p. 377-388, 2015.

Automated histological concept identification

In these set of publications, the ability of multi-scale descriptors and graph based features to capture the differences in histopathological concepts from skin whole slide images was investigated. Specifically, strategies to distinguish between normal histological structures and basal cell carcinoma were explored [55, 57, 58].

- **David E. Romo**, Juan D. García-Arteaga, Pablo Arbeláez and Eduardo Romero. *A Discriminant Multi-Scale Histopathology Descriptor using Dictionary Learning*, Proceedings of the SPIE Medical Imaging, 2014.

- **David Romo-Bucheli**, Ricardo Moncayo, Angel Cruz-Roa and Eduardo Romero. *Identifying histological concepts on basal cell carcinoma images using nuclei based sampling and multi-scale descriptors*, Proceedings of the 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York, 2015, pp. 1008-1011.

- **David Romo-Bucheli**, Germán Corredor, Juan D. García-Arteaga, Viviana Arias, and Eduardo Romero. *Nuclei graph local features for basal cell carcinoma classification in whole slide images.*, Proceedings of the 12th International Symposium on Medical Information Processing and Analysis (SIPAIM), pp. 101600Q-101600Q, 2016.

Cancer grading - comparison with manual grading

In this publication a multi-scale descriptor is used to characterize the nuclear pleomorphism grading of breast cancer images [50]. The construction of a nuclei dictionary using the multi-scale descriptor enables a bag of feature strategy to correctly highlight differences between microscopical fields labelled with different nuclear grades. The proposed method achieved a second place in the atypia classification task at the mitos-atypia 2014 challenge [1].

[1] https://mitos-atypia-14.grand-challenge.org/results2

- Ricardo Moncayo, **David Romo-Bucheli**, and Eduardo Romero. *A Grading Strategy for Nuclear Pleomorphism in Histopathological Breast Cancer Images Using a Bag of Features (BOF)*, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, LNCS, 2015.

Automated pathological markers and outcome - comparison with genomic tests

In these set of publications, the correlation between automated pathological markers (such as mitotic activity and normal gland formation within cancerous regions) and cancer risk as asserted by genomic test were investigated[61, 59, 60]. In a large set of high power fields extracted from a population of estrogen receptor positive breast cancer, the pathological markers were quantified and the correlation with the cancer risk was evaluated.

- **David Romo-Bucheli**, Andrew Janowczyk, Eduardo Romero, Hannah Gilmore, Anant Madabhushi. *Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images* , Proceedings of the SPIE Medical Imaging, 2016.

- **David Romo-Bucheli**, Andrew Janowczyk, Eduardo Romero, Hannah Gilmore, Anant Madabhushi. *Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images*, Nature Publishing Group, Scientific Reports, vol. 6, no. 32706, 2016.

- **David Romo-Bucheli**, Andrew Janowczyk, Eduardo Romero, Hannah Gilmore, Anant Madabhushi. *A deep learning based strategy for ideintifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers*, Cytometry Part A, 2017. doi:10.1002/cyto.a23065

Thesis outline

The remaining chapters of the thesis are organized as follows:

- **Chapter 2: *Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images***

  This chapter presents a comparison between the automated tubule quantification and the cancer risk in breast cancer. A deep learning model was used to identify nuclei belonging to a tubular structure. After this, the model is used to quantify the tubule formation in groups of low, intermediate and high risk breast cancer. The resulting measurements are correlated with the different cancer risk groups.

- **Chapter 3:** ***A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in Estrogen Receptor Positive Breast Cancers*** In this chapter a deep learning model to identify mitosis is presented. The model is used to quantify the mitotic activity in a population of breast cancer specimens that have a high, intermediate and low risk of recurrence, as asserted by the Oncotype DX genome test. The quantified mitotic activity is compared with the risk groups.
- **Chapter 4:** ***Conclusions*** The final chapter in this work presents some conclusions and discuss the potential impact of automated pathological markers in histological image analysis. Also, future research directions and perspectives are discussed.

2. Automated Tubule Nuclei Quantification and Correlation with Oncotype DX risk categories in ER+ Breast Cancer Whole Slide Images

Early stage estrogen receptor positive (ER+) breast cancer (BCa) treatment is based on the presumed aggressiveness and likelihood of cancer recurrence. Oncotype DX (ODX) and other gene expression tests have allowed for distinguishing the more aggressive ER+ BCa requiring adjuvant chemotherapy from the less aggressive cancers benefiting from hormonal therapy alone. However these tests are expensive, tissue destructive and require specialized facilities. Interestingly BCa grade has been shown to be correlated with the ODX risk score. Unfortunately Bloom-Richardson (BR) grade determined by pathologists can be variable. A constituent category in BR grading is tubule formation. This study aims to develop a deep learning classifier to automatically identify tubule nuclei from whole slide images (WSI) of ER+ BCa, the hypothesis being that the ratio of tubule nuclei to overall number of nuclei (a tubule formation indicator - TFI) correlates with the corresponding ODX risk categories. This correlation was assessed in 7513 fields extracted from 174 WSI. The results suggests that low ODX/BR cases have a larger TFI than high ODX/BR cases ($p < 0{,}01$). The low ODX/BR cases also presented a larger TFI than that obtained for the rest of cases ($p < 0{,}05$). Finally, the high ODX/BR cases have a significantly smaller TFI than that obtained for the rest of cases ($p < 0{,}01$). *The complete content of this chapter was published in Scientific Reports Journal [59].*

2.1. Introduction

The primary conundrum in treatment and management of early stage estrogen receptor positive (ER+) breast cancer (BCa) is identifying which of these cancers are candidates for adjuvant chemotherapy and which patients will respond to hormonal therapy alone. ODX and other gene expression tests have allowed for distinguishing the more aggressive ER+ BCa requiring adjuvant chemotherapy from the less aggressive cancer benefiting from hormonal therapy alone. However these gene expression tests tend to be expensive, tissue

destructive and require physical shipping of tissue blocks for the test to be done. Interestingly BCa grade in these tumors has been shown to be highly correlated with the ODX risk score [38, 65, 1]. Unfortunately studies have shown that Bloom Richarsdon (BR) grade determined by pathologists can be highly variable [21]. The three constituent categories within the BR grading system are mitotic index, tubule formation and nuclear pleomorphism. Tubule formation is defined as the percentage of cancer tissue that still contains normal tubules. According to Elston and Ellis guidelines [26], tumor cell clusters with glandular formation are also counted (Figure **2-1** shows some examples of tubule delineations for low and high risk BCa). Tubule scoring is determined by estimating tubule area and assigning to one of three categories: (i) $> 75\,\%$, (ii) between $10\,\% - 75\,\%$, and (iii) $< 10\,\%$. However, this estimation is highly influenced by experience of the pathologist. Additionally, previous studies have shown the correlation between manually determined tubule score and ER+ breast cancer prognosis and ODX risk categories[29, 39].

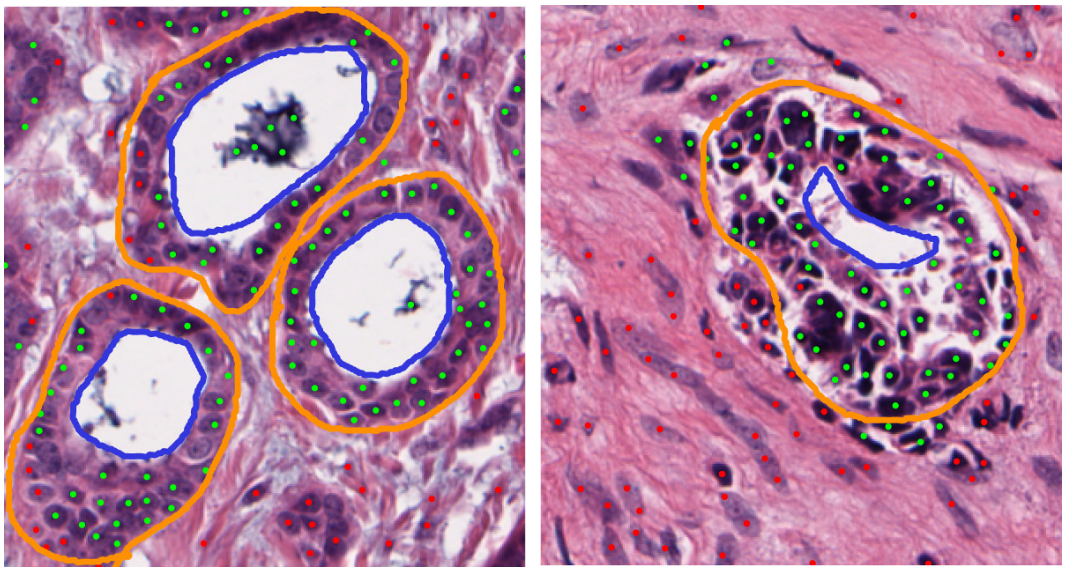

Figure 2-1.: Breast Cancer tissue showing (left) high and (right) low tubule formation. Lumen is delineated by blue lines. Tubules are delineated by orange lines, where nuclei inside these boundaries represent the tubule nuclei used in our approach. Green and red dots correspond to nuclei candidates classified as tubule or non-tubule nuclei by our DNN classifier. In the high ODX (right) image, the cells have lost their capacity to form tubules with a rounded lumen.

.

Since histologic criteria (such as tubule, nuclei pleomorphism, and mitotic activity) are used in pathological grading systems, several works using automated extraction algorithms have been proposed to quantify such criteria [67]. Tubule detection has been previously addressed in the literature [20, 7, 48]. Typically these approaches focus on the identification of tubule lumen (see Figure **2-1**). Strategies focused on identifying tubules based off the lumen present

a couple of challenges. Firstly the shape and size variability of the gland lumen makes accurate modeling of the tubules difficult. Secondly several structures, besides tubules, also contain lumen, i.e., blood vessels and other types of glands. Morphological operators have been used to connect proximal cancerous cells and generate blob structures [20]. These blobs were identified as tubules when they were found to be surrounding a white space or lumen. Another approach using the O'Callaghan neighborhood graph to impose structural constraints on lumen, allowed for identification of true lumen with an accuracy of 86 % [7]. An accuracy of 89 % was obtained in the classification task of low (tubular BR score 2 and 3) and high tubule formation (tubular BR score 1). A similar strategy, using k-means to identify lumen followed by a level set based segmentation approach enabled the identification of the surrounding nuclei layer [48].

The deep neural network (DNN) is a deep learning architecture that comprises more than two hidden layers. In supervised classification settings, a DNN uses the backpropagation algorithm to update its internal weights according to the label of input exemplars [41]. Some applications of the DNNs in histological image analysis include the mitosis identification task [16] and the localization of regions of interest in histological images [19].

With the recent emergence of whole slide tissue scanning and digital pathology [47, 10, 31] there has been substantial interest in developing automated computerized histologic predictors of tumor grade and outcome for several diseases including oropharyngeal squamous cell carcinoma [46], prostate cancer [45, 44] and glioblastoma [14]. The correlation of computerized extracted features with breast cancer survival has also been explored. Beck et al. [8] performed a comprehensive analysis of several automatically quantified morphological features and their relationship with breast cancer survival. The authors reported a strong association of automatically extracted stromal features with survival in a set of 576 H&E breast cancer tissue microarray (TMA) images. Tambasco et al. [64] used fractal analysis to compute the morphological complexity of 379 pan-cytokeratin stained TMA images. A significant association of survival with the computed fractal dimension was found. The correlation of automated extracted features with Oncotype DX risk score and risk categories has been investigated in a couple of studies. Basavanhally et al. [6] showed that nuclear graphs built using Delaunay triangulation and minimum spanning trees can be used to distinguish breast cancer images with low and high recurrence ODX scores (RS). The authors used 37 H & E stained images from a cohort of 17 patients at 20× magnification and obtained a mean accuracy of 84,15 % in distinguishing samples with low and high RS. Also, the combination of computer extracted features from both H&E and CD34 IHC stained images in a cohort of 29 patients (9 with low RS, 11 intermediate RS and 9 with high RS) [5] was shown to distinguish high and low ODX risk patients. The authors reported an average classifier accuracy of 91 % for distinguishing high and low RS cases. Other studies have explored the association between manually identified pathological measurements (e.g. nuclei grade, mitotic index, tubule degree) and the Oncotype DX score. Both Flanagan et al. [29] and Klein et al. [39] used regression analysis to obtain a set of equations that predicts Oncotype DX score based

on histological variables such as nuclei grade, mitotic index, tubule formation degree among others. After eliminating cases from the intermediate risk category, concordance between the ODX score and the estimated score (using the obtained equations) range from 96,9 % to 100 %.

The contributions of the work presented in this paper are twofold. Firstly we aim to evaluate a customized DNN for automatic quantification of tubules in whole slide images (WSI). Secondly we seek to evaluate whether tubule score automatically identified by the DNN is correlated with the risk categories determined by ODX in a cohort of 174 patients. Our approach comprises the following main steps. First, a blue ratio transform is used to detect nuclei candidates. Image patches, each containing a nucleus, are then extracted. These patches are manually labeled as containing a tubule or not. The patches are used to train a DNN classifier to identify tubule nuclei in WSI. After tubule nuclei identification, the ratio between tubule nuclei and overall number of nuclei is computed as a tubule formation indicator (TFI).

The rest of this paper is organized as follows: Section 2 describes the methodology used for training and testing the DNN tubule nuclei classifier. Section 3 presents the experimental design to study the correlation of the TFI with ODX risk categories. Section 4 describes the results of the statistical experiments and the distribution of the TFI for the ER+ BCa cases. Finally, in Section 5 we present the main conclusions of our work.

2.2. Methodology

The whole methodology to use the automated TFI to study its correlation with ODX score and BR grading in WSI is presented in Figure **3-1**.

2.2.1. Nuclei detection

First, an automated algorithm based on blue ratio transformation[15] is used to detect nuclei. After computing the blue ratio transform, a global threshold computed by using Otsu's method [52] is used to obtain a binary image. Then, an opening operation is applied. The centroid of each connected component corresponds to the centroid of a nucleus candidate. The nuclei detection algorithm is a lightweight method that provides a nuclei rough estimation that was found to be representative of the true nuclei population in terms of the TFI, as shown by the experiments described in the supplementary information (See Appendix G).

2.2.2. Curating the Learning Set

An RGB patch is extracted (size 64×64 at $20\times$ magnification with a spatial resolution of approximately $0{,}5\mu m$ per pixel) around the centroid of each candidate nuclei. This patch

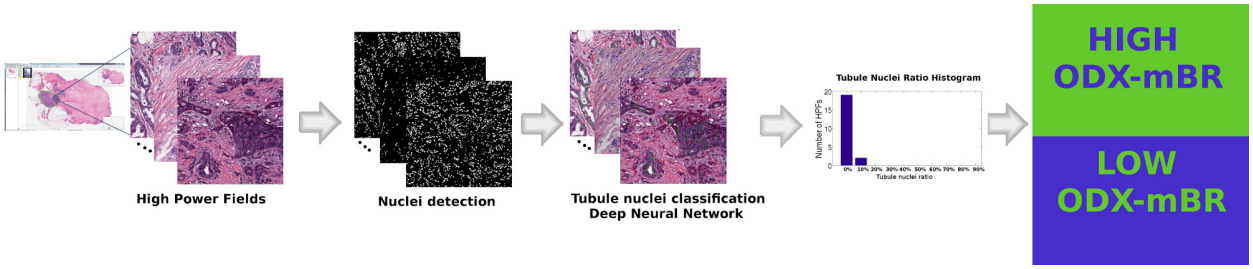

Figure 2-2.: Overall diagram flow showing the steps to analyze the correlation of the tubule formation indicator with ODX score and BR grade. Several high power fields from a whole slide images are extracted. A nuclei detection method is then applied on each high power field. Each of the candidate nuclei is classified as tubule or not using a DNN classifier. Subsequently, the mean tubule nuclei ratio to total number of nuclei per high power field for each whole slide image is computed and analyzed with respect to the corresponding ODX risk category and BR grade.

is labeled as either tubule or not, according to an annotation supplied by an expert pathologist (The expert breast pathologist annotation corresponds to a manual delineation of each tubule). These pathologist annotated patches are then used to train the DNN classifier. Exemplar RGB patches belonging to the tubule class and non-tubule class are presented in Figure **2-3**.

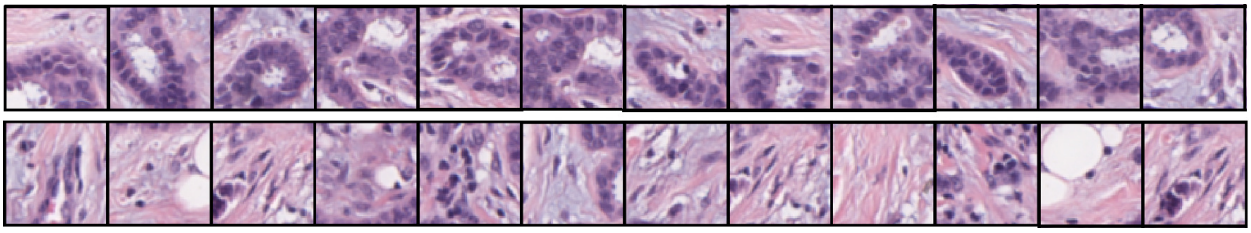

Figure 2-3.: Examples of image patches used for training. **Top Row:** The tubule class. **Bottom row:** The non-tubule class. Each patch center corresponds to a nucleus candidate centroid.

The DNN architecture is illustrated in Figure **3-2** and is composed of three blocks: a convolution neural network (CNN), a Rectifier Linear Unit (ReLU) and a maximum pool (max pool) operator. Finally, two fully connected layers yield the probability representing the membership of the nucleus to the tubule class.

2.2.3. Independent testing of the DNN classifier

During testing, the nuclei detection algorithm is used to identify candidate nuclear centroids. These patches then fed to the DNN, as shown in Figure **3-2**. This process enables the generation of their tubule class membership probability. If the probability is higher than 0,5, the patch is assigned to the tubule class.

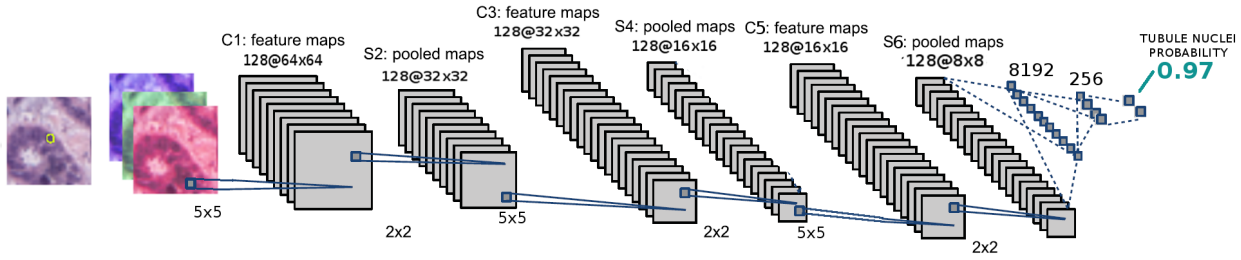

Figure 2-4.: Deep learning architecture used to classify nuclei. A patch containing a nucleus feeds the deep neural network. The probability of the nucleus being part of a tubule is based on the output of the deep neural network classifier.

The DNN performance was evaluated on a dataset with 61 high power fields that were extracted from 11 WSI. Whole tubule structures (including epidermis surrounding the lumen) had been previously annotated by an expert pathologist. A 5-fold cross validation setup was used, ensuring each fold was split at the patient level.

Evaluation measures (F_{score}, precision, recall (sensitivity) and specificity for the tubule nuclei class[53]) were computed for each of the 5-folds. The average +/- standard deviation of the F_{score}, precision, recall and specificity were: $0{,}59 \pm 0{,}14$, $0{,}72 \pm 0{,}12$, $0{,}56 \pm 0{,}2$ and $0{,}9 \pm 0{,}06$ respectively (see Figure **2-5**).

Observe that the recall for the tubule identification is lower than the specificity, indicating that a classification error is more likely for a tubule nuclei than for a non-tubule nuclei. Also, the variability of tubule sizes and shapes may explain the higher standard deviation obtained with the recall measure. Detailed results for each fold are presented in Table **2-1**.

The detection results in Table **2-1** suggest that the tubule detector has a high specificity, a finding that might be caused by the unbalanced nature of the problem (there is a larger number of non-tubule nuclei as opposed to tubule nuclei in the BCa specimens). Also the tubule nuclei exhibit a substantially large inter-subject variation. The tubule nuclei samples used during training might not be adequate to capture all the variability observed in tubules from different patients.

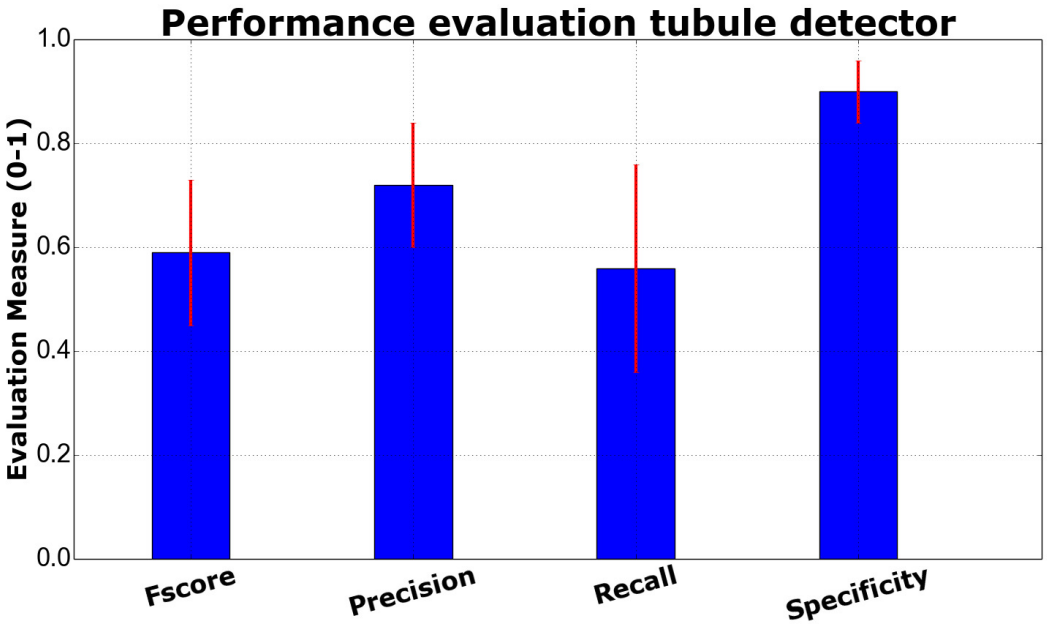

Figure 2-5.: Performance evaluation measures for the tubule nuclei detection task in a 5-Fold cross validation setup and involving images extracted from $N = 11$ patients.

2.3. Experimental Design

2.3.1. Data Description

A set of WSI extracted from 174 patients with ER+ BCa were used in this study. At most 50 high power fields per WSI were selected: the selected high power fields were those with the lower number of tubule nuclei ratio. This selection avoids high power fields with unusually large number of detected tubule nuclei (outliers). All of these high power fields were sampled from cancerous regions previously identified by an expert pathologist.

2.3.2. Correlation with ODX risk groups via t-test analyses

After identifying the tubule nuclei the TFI was computed: the ratio between the tubule nuclei and the total number of nuclei. This TFI is evaluated as a potential risk predictor.
In order to compare the TFI with the risk associated to each BCa sample, the set was divided into a) High, b) Intermediate and c) Low risk categories according to the ODX score. Additionally, the BR grade is also used to define: d) The high ODX-high grade group (with both high ODX and BR score-HH), e) The low ODX-low grade group (with both low ODX and BR score-LL), f) All the BCa cases that don't belong to the HH group (HHc

	F-score	*Precision*	*Recall (Sensitivity)*	*Specificity*
Fold 1	0,34	0,89	0,21	0,98
Fold 2	0,7	0,73	0,67	0,9
Fold 3	0,71	0,81	0,63	0,93
Fold 4	0,64	0,59	0,7	0,84
Fold 5	0,59	0,59	0,59	0,83
Average	$0{,}59 \pm 0{,}14$	$0{,}72 \pm 0{,}12$	$0{,}56 \pm 0{,}2$	$0{,}9 \pm 0{,}06$

Table 2-1.: 5-Fold validation results for the tubule detection across $N = 11$ patients. The F-score, precision, recall and specificity for the tubule classification are shown. The final row presents the average and standard deviation for each of the four measures across all $N = 11$ patients and across the 5 folds.

group) and g) All the BCa cases that don't belong to the LL group (LLc group). The dataset categorization is indicated in Table **2-2**.

The t-test statistical analysis was applied to compare the distribution of the automated TFI with the high, intermediate and low ODX risk groups as well as the BCa cases with both a high ODX score and high grade and also cases with both low ODx score and low BR grade. The t-test for all the experiments was performed with equal mean and unequal variance hypothesis. Specifically, the t-test was applied to compare the different groups as described below:

- The high ODX group against the low ODX group
- The high ODX group against both the intermediate and low ODX group
- The low ODX group against both the high and intermediate ODX group
- The high ODX-high grade (HH Group) against the low ODX-low grade (LL group)
- The high ODX-high grade (HH Group) against all the other cases (HHc group) and
- The low ODX-high grade (LL group) against all the other cases (LLc group)

2.3.3. Correlation with ODX risk groups via ROC analysis

The risk prediction capability of the TFI was also evaluated using a Receiver Operating Curve (ROC). For doing so, the binary classification task was based solely in the tubule nuclei ratio: each WSI with a mean tubule ratio above a particular threshold is classified as low ODX. By varying the threshold from $[0, 1]$ is possible to generate the ROC curve. In this particular experiment the goal was to distinguish the HH and LL categories (see Table **2-2**).

BCa Groups	*Description*	*Number of cases*
High ODX	$ODX > 30$	24
Low ODX	$ODX < 18$	95
Intermediate ODX	$18 \leq ODX \leq 30$	55
High ODX-high grade (HH)	Both $ODX > 30$ and $BR > 7$	15
Low ODX-low grade (LL)	Both $ODX < 18$ and $BR < 6$	42
HHc group	All BCa cases that do not belong to HH group	159
LLc group	All BCa cases that do not belong to LL group	132

Table 2-2.: ODX score and BR grading rules used to split the dataset into high, intermediate and low ODX categories. The corresponding number of cases for each group is presented in the last column.

2.4. Results

2.4.1. Correlation with ODX and BR risk categories via t-test analyses

The DNN classifier was applied to the 174 WSI previously described. Qualitative results for high, intermediate and low ODX cases can be seen in the Figure **2-6**. The significant t-test results for the comparison between the risk groups is presented in Table **3-1**.

Risk group comparison	**p-values (Unequal variance)**
H vs L	0.1633
H vs L and I	0.2731
H and I vs L	0.0998
HH vs LL	**0.0021**
LLc vs LL	**0.0145**
HH vs HHc	**0.0097**

Table 2-3.: Statistical comparison of the deep learning tubule classifier in distinguishing different risk groups. Note that statistically significant differences were only observed for 3 of the 6 comparative experiments performed.

When observing the group distribution according to ODX score, it is difficult to distinguish between low and high ODX groups. However, when combined ODX and BR groups are analyzed, the high and low risk groups show different distributions as shown in Figure **2-7**. Results in Figure **2-7** reveal that the automated TFI is significantly different for the groups that have low ODX-low grade and high ODX-high grade. The HH group had a mean tubule nuclei ratio per high power field of 0,029. In contrast, the LL group had a mean tubule nuclei ratio of 0,126. The two groups are significantly different ($p < 0{,}01$ with 95 % CI [0,04, 0,16]). The differences in the TFI is still significant when we compare the HH group against the BCa cases that did not belong to this group ($p < 0{,}01$ with 95 % CI$[0{,}013, 0{,}085]$). The mean for non HH cases was 0,078. Finally, the difference in the average TFI value was also significant when comparing the LL group with the BCa cases outside this group ($p < 0{,}05$

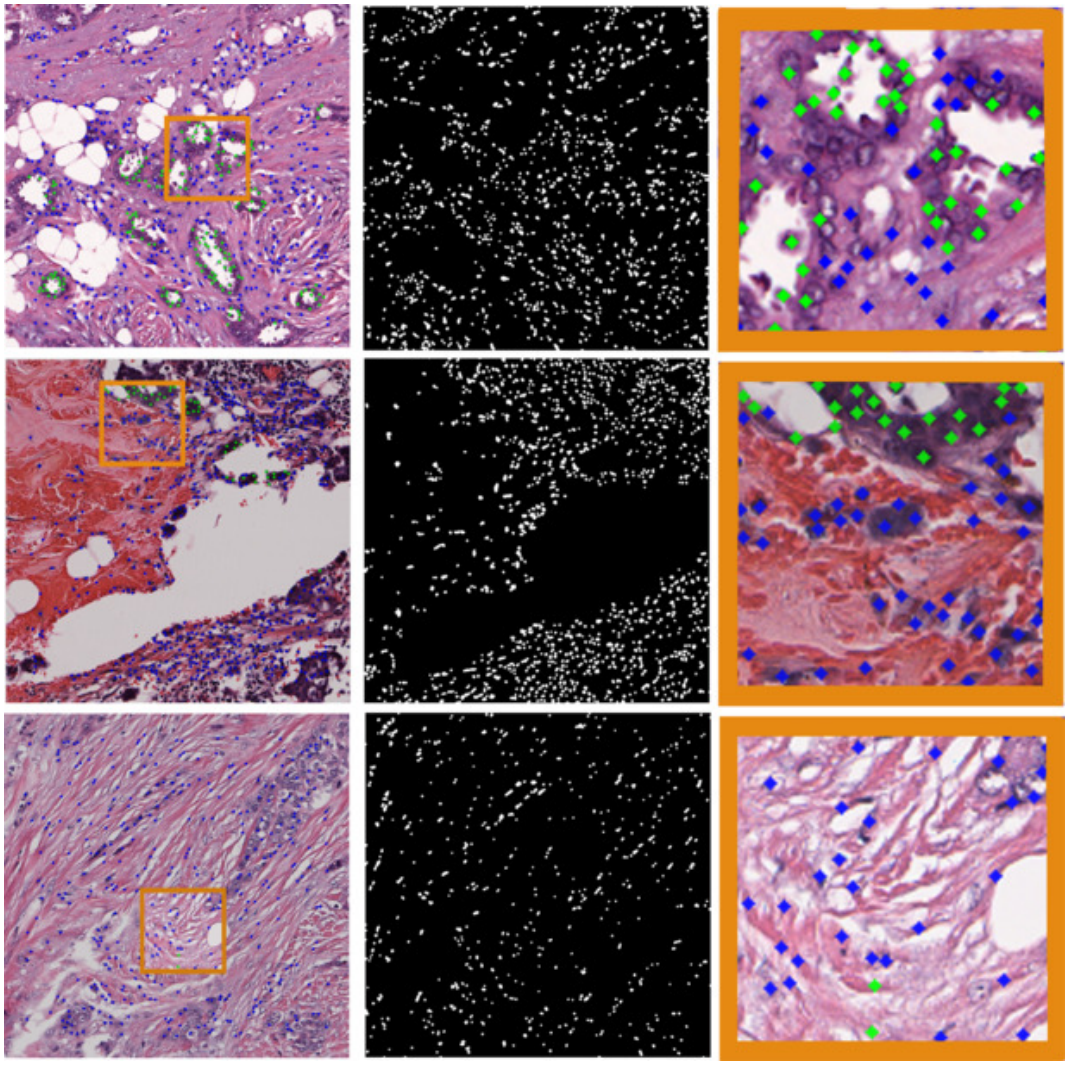

Figure 2-6.: Tubule nuclei identification process for high power fields extracted from low ODX (top row), intermediate ODX (middle row) and high ODX (bottom row) breast cancers. In the first column, the high power field at $\times 20$ magnification is depicted. In the second column, the resulting mask showing the nuclei centroids after the nuclei detection process is presented. The third column shows the DNN classification of each nucleus either as a tubule nucleus (green dot) or a non-tubule nucleus (blue dots). Each image in the right column corresponds to a close up in the selected region (orange rectangle) depicted in the left most column. For the low ODX high power field, a significant number of tubule nuclei are identified. Observe also that some false negatives are not uncommon in the nuclei surrounding the tubule lumen. On the other hand, the high and intermediate cases have a substantially lower number of tubule nuclei. Some false positive (false tubule nuclei) errors are also visible in the right most column.

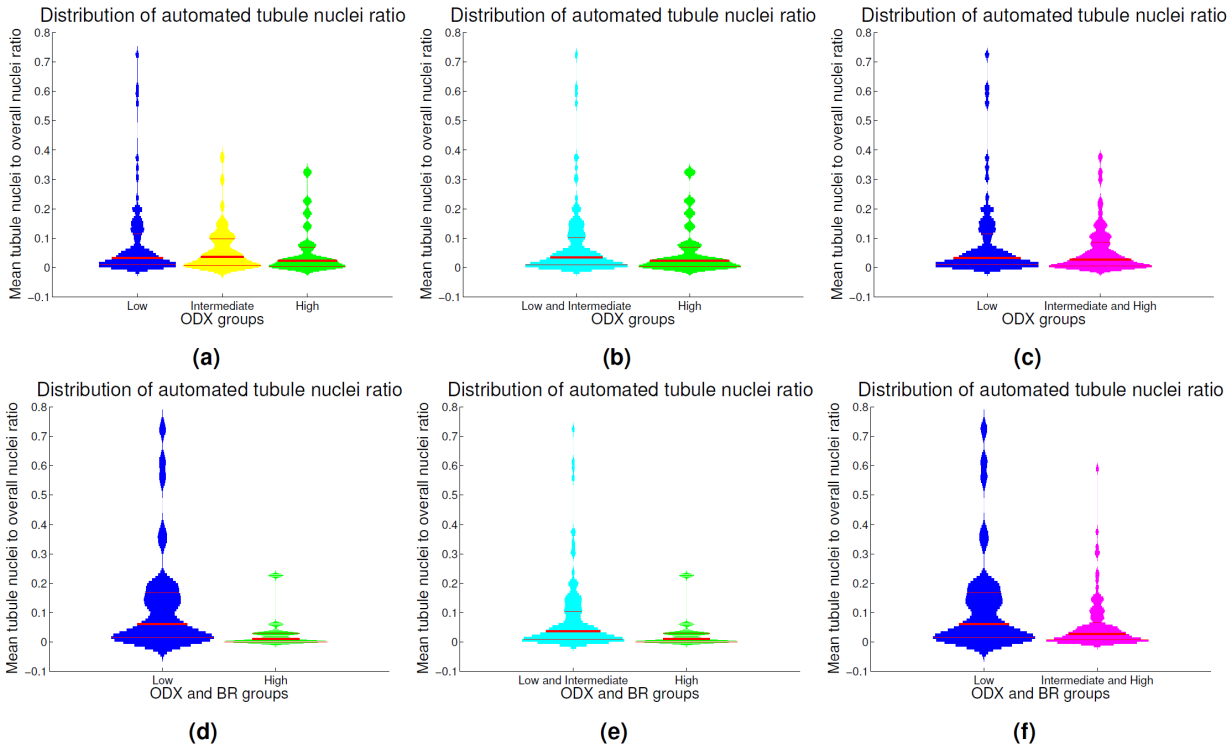

Figure 2-7.: Violin plots depicting the mean tubule nuclei ratio in high power fields extracted from the different ODX risk groups. The histogram associated to each violin plot is smoothed using a normal kernel. Red lines in the violin plot show the location of the lower quartile (q_1), the median and the upper quartile (q_3). Low (blue), intermediate (yellow) and high (green) ODX groups are shown in the top row (**a**). The distribution of the low and intermediate groups (cyan) against the high ODX group is presented in (**b**). The low group against the intermediate and high ODX groups (magenta) are presented in (**c**). The distribution for the groups with low ODX-low grade and high ODX-high grade are depicted in (**d**). High ODX-high grade against all the other BCa cases and low ODX-low grade against all the other BCa cases are presented in (**e**) and (**f**) respectively.

with 95 % CI[0,014, 0,12]). The cases that did not belong to the LL group had a mean tubule nuclei ratio of 0,057.

2.4.2. Correlation with ODX and BR risk categories via ROC curve

The distribution of the histologic images (ODX score vs tubule nuclei ratio) for the HH and LL groups is shown in the left column of Figure **2-8**. While a low mean tubule nuclei ratio appears to require additional analysis to determine its risk category, it is observed that a WSI with a high tubule nuclei ratio is very likely to be member of the low ODX risk category.

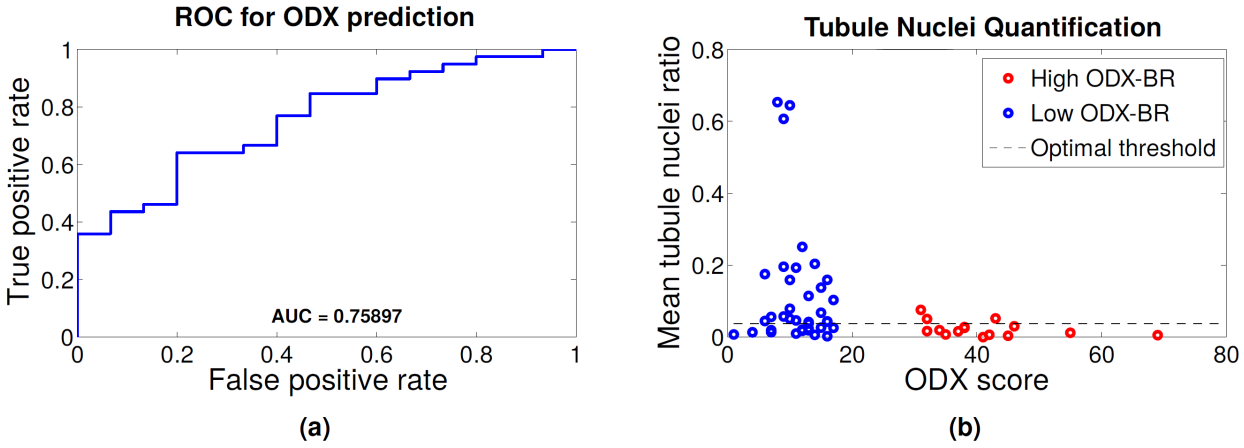

Figure 2-8.: (**a**) Receiver operating characteristic (ROC) curve for the prediction of low ODX using only the tubule nuclei ratio feature. (**b**) Mean automated tubule nuclei ratio for each whole slide image. The high (red) and low (blue) ODX score groups are depicted. The x-axis represents the underlying ODX score of each sample. The y-axis represents the tubule nuclei ratio. Observe that the high ODX image have a low tubular density. A high tubule nuclei ratio is very likely associated with a low ODX image. Optimal threshold obtained for the ROC curve (threshold at which the ROC curve is closest to point $[0, 1]$) is also shown.

The Receiver Operating Curve (ROC) for the binary classification task using only mean tubule nuclei ratio for each WSI is presented in the right column of Figure **2-8**. The WSI with a mean tubule ratio above the threshold is classified as low ODX. The ROC curve shows that the tubule nuclei ratio yields an area under the curve (AUC) of 0,76 in distinguishing the low ODX-low grade from the high ODx-high grade categories.

2.5. Concluding Remarks

In this paper we rigorously investigated the problem of objectively computing the tubule nuclei ratio, a potential computational histologic image biomarker of disease risk and aggressiveness in ER+ BCa. To evaluate whether automatically TFI was associated with the risk category determined by the Oncotype DX test, a deep learning classifier was developed to automatically identify tubules based off the surrounding nuclei. The automatically determined TFI was then evaluated in terms of its ability to distinguish the low and high ODX risk categories and cases with different permutations of ODX risk and grade. On a cohort of 174 WSI, the TFI was found to be significantly different for the BCa cases with low ODX-low grade and high ODX-high grade. When comparing the high ODX-high grade group with all the other BCa cases, the TFI was still significantly lower. Likewise, the calculated tubule quantification measure was larger in the BCa cases with low ODX-low grade compared to the remaining BCa cases.

The automated TFI appears to have a slightly weaker correlation with ODX risk categories than other previously investigated computerized image features such as nuclear architecture [6]. However it has been previously shown that using a combination of automated features (even extracted from differently stained samples from the same patient), might increase the ability to predict the corresponding ODX risk category [5]. Hence, developing strategies to integrate information from predictors that use different histological features (e.g. nuclear architecture, mitotic count, tubule density) will be a future research endeavor.

Automated tubule quantification could be potentially useful in streamlining clinical pathology workflows. The automated quantification aims to standardize the breast cancer grading and risk assessment process and reduce inter-reader variability. Our newly presented method was evaluated within manually selected cancerous regions. However, automatic delineation of regions of diagnostic interest is an open research problem [19]. Future work will focus on improving the tubule detector performance, validating our approach on larger test cohorts and incorporating automatic region of interest selection methods.

3. A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in Estrogen Receptor Positive Breast Cancers

The treatment and management of early stage estrogen receptor positive (ER+) breast cancer is hindered by the difficulty in identifying patients who require adjuvant chemotherapy in contrast to those that will respond to hormonal therapy. To distinguish between the more and less aggressive breast tumors, which is a fundamental criterion for the selection of an appropriate treatment plan, Oncotype DX (ODX) and other gene expression tests are typically employed. While informative, these gene expression tests are expensive, tissue destructive, and require specialized facilities. Bloom-Richardson (BR) grade, the common scheme employed in breast cancer grading, has been shown to be correlated with the Oncotype DX risk score. Unfortunately, studies have also shown that the BR grade determined experiences notable inter-observer variability. One of the constituent categories in BR grading is the mitotic index. The goal of this study was to develop a deep learning (DL) classifier to identify mitotic figures from whole slides images of ER+ breast cancer, the hypothesis being that the number of mitoses identified by the DL classifier would correlate with the corresponding Oncotype DX risk categories. The mitosis detector yielded an average F-score of 0,556 in the AMIDA mitosis dataset using a 6-fold validation setup. For a cohort of 174 whole slide images with early stage ER+ breast cancer for which the corresponding Oncotype DX score was available, the distributions of the number of mitoses identified by the DL classifier was found to be significantly different between the high vs low Oncotype DX risk groups ($p < 0{,}01$). Comparisons of other risk groups, using both ODX score and histological grade, were also found to present significantly different automated mitoses distributions. Additionally, a support vector machine classifier trained to separate low/high Oncotype DX risk categories using the mitotic count determined by the DL classifier yielded a 83,19 % classification accuracy. *The complete content of this chapter was published in Cytometry Part A journal [60].*

3.1. Background

Modern treatment of early stage estrogen receptor positive (ER+) breast cancer requires a precise identification of which cases will benefit from additional adjuvant chemotherapy versus those indicating solely hormonal therapy. A distinction between the more and less aggressive breast tumors, required for planning the treatment, is usually performed with Oncotype DX (ODX) and other gene expression tests. In general, the more aggressive cancers require adjuvant chemotherapy while the more benign respond well to hormonal therapy alone. However these gene expression tests tend to be expensive, tissue destructive, and require specialized facilities for processing.

The Oncotype DX risk score has been demonstrated to be highly correlated with breast cancer grade [38, 65, 1]. Unfortunately, the grades yielded by the standard Bloom-Richardson system (BR) have been found to be highly variable as a result of both observer experience and tumor presentation[21]. The BR consists of assigning a value to stratify three properties: (a) the nuclear pleomorphism score aims to characterize the variance in nuclear size and appearance, (b) the tubule density score is intended to reflect the percentage of cancer tissue that contains normal tubules, and (c) the mitotic index aims to quantify the number of mitoses in a specific number of high power fields (HPFs; typically 10)[27]. Although the number of mitoses is regarded as an important prognostic indicator, the overall predictive power is limited by significant inter-reader variability due to potentially differently selected fields[49]. In spite of these challenges, studies have shown that Oncotype DX is highly correlated with the mitotic grade for ER+ breast cancers[29].

The recent addition of whole slide imaging capabilities to pathology has sparked notable interest in the use of automated computerized histologic predictors of tumor grade and outcome [46, 45, 44, 14]. In the context of breast cancer, Basavanhally et al. [6] showed that nuclear graphs built using Delaunay Triangulation and Minimum Spanning Trees can be used for distinguishing low and high recurrence score breast cancer cases.

Of late, there has been substantial interest in automating the process of identifying mitotic figures in whole slide pathology images [67]. Larsen et al. [43] used color intensity histograms, gradient orientation histogram and shape index histograms to identify mitoses. Wang et al. [69] proposed to use a convolution neural network and a set of handcrafted features combined with a random forest classifier to identify mitotic figures in WSI. Janowczyk et al. [37] used deep learning approaches to perform several histologic image analysis tasks, including mitotic identification. Veta et al. [66] evaluated the performance of several state-of-the-art mitotic detection methods and found that a Deep Neural Network (DNN) yielded the best overall performance with an overall F-score of 0,61 [16] in the test set of the AMIDA2013 challenge [66].

A neural network architecture consisting of more than two layers is commonly referred as a deep neural network. In supervised classification settings, a DNN uses the backpropagation algorithm to update its internal weights according to the label of input exemplars [41]. The

DNN for mitotic detection is usually trained using a set of image patches that are within a defined d pixel radius to a mitotic nuclei [16].
The contributions of the work presented in this chapter are twofold. Firstly we aim to evaluate a customized DNN for automatic quantification of mitosis in WSI. Secondly, we seek to evaluate whether the automated mitotic index identified by the DNN correlates with the risk categories determined by ODX.

3.2. Experimental Methods

The methodology used to automate mitotic nuclei identification and classify WSI into risk BCa groups as defined by both ODX score and BR grading is illustrated in Figure **3-1**.

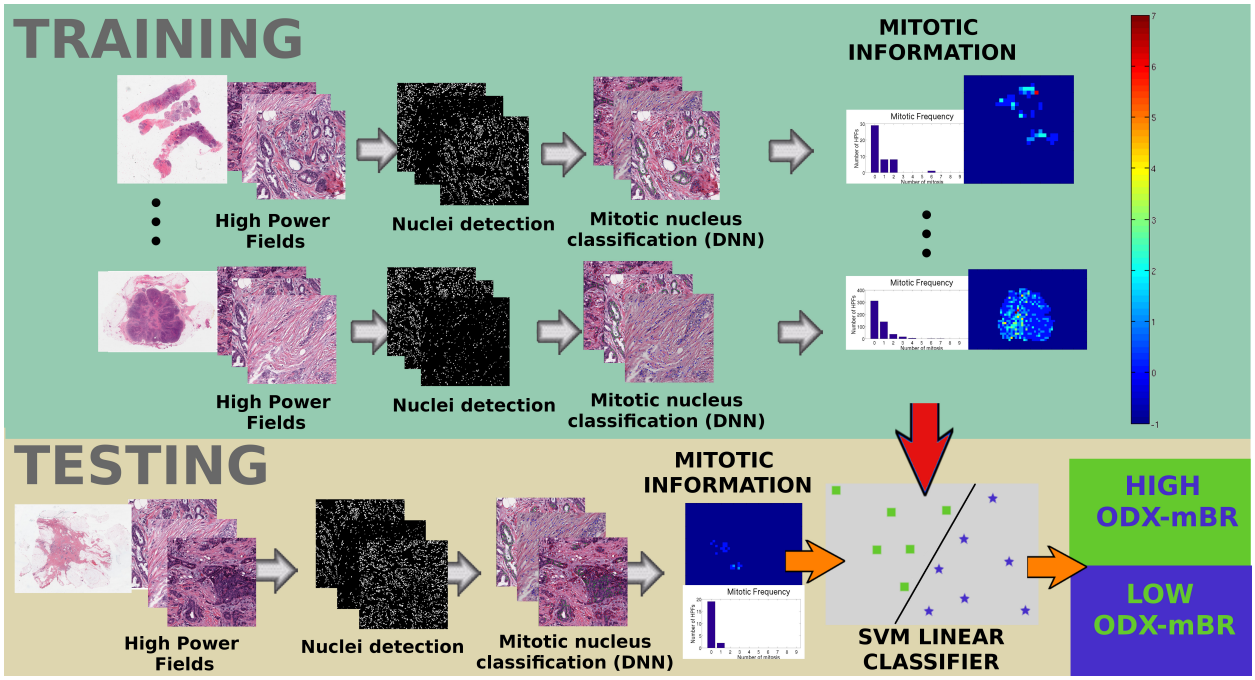

Figure 3-1.: Overall diagram flow showing the steps to analyze the correlation of the automated mitotic count with ODX and to use mitotic information in BCa risk stratification. In the training stage, high power fields (HPFs) from several WSI are extracted and a nuclei detection method is applied on each high power field. Each of the candidate nuclei is classified as mitotic or not using a DNN classifier. Subsequently, the mitotic information is used to train a linear support vector machine classifier. At testing stage, nuclei identification algorithm and the DNN classifier are used again in the cancerous regions of the WSI. Finally, the resulting mitotic information is used by the support vector machine classifier to predict the WSI risk either as high or low.

3.2.1. DNN for mitotic identification

The mitosis detection module involved the training of a DNN model with Hematoxylin and Eosin stained images. Candidate nuclei were found by an automated nuclei detection algorithm at 40X magnification. Candidate nuclei, separated by less than 20 pixels from the expert annotations, were labeled as mitotic. The approach starts by mapping each RGB point to a grayscale intensity value [15] using: $Br = \frac{100B}{1+R+G} * \frac{256}{1+B+R+G}$ (blue ratio transform). The resulting image highlights the chromatin information (observable in mitotic nuclei) and an adaptive threshold converts the blue ratio image into a binary map, using the gray-level histogram information. The optimal threshold is supposed to maximize the separability of two resulting pixel clusters [52]. Subsequently, a set of morphological operations (closing and opening) further improve the selection of the candidate nuclei by removing noise. The connected components are finally selected as candidate nuclei. The centroid of each candidate nucleus corresponds to the center of the associated RGB color patch (size 64×64 at $40\times$ magnification) which is then fed into the DL classifier to determine whether or not it is a mitotic nucleus.

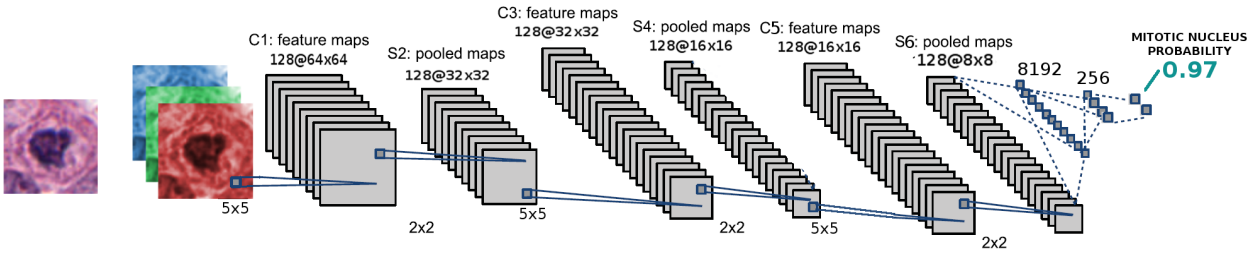

Figure 3-2.: Deep learning architecture for mitosis detection. A 64×64 size patch is extracted at $40\times$ magnification. The patch contains a candidate nucleus which is fed to the DNN. The DNN is composed of several convolutional neural networks, each followed by non-linear and pooling operations. At the end, two fully connected layers are used to estimate the probability of a patch containing a mitotic figure. The probability that the image patch contains a mitotic nucleus is then assigned by the classifier.

The DNN architecture in Figure **3-2** consists of three identical blocks, each composed of a convolution neural network (CNN), a Batch normalization layer, a Rectifier Linear Unit (ReLU) and a maximum pool (max pool) operator. Each CNN layer corresponds to a set of convolutional filters learned from the mitotic and non-mitotic classes. The Batch normalization layer imposes a transformation of the neural weights to solve the internal covariate shift problem[35]. In practice, this normalization layer allows the method to achieve improved learning rates during the DNN training stage. The proposed network architecture reduces the input patch to an output signal which coincides with the probability that the input contains a mitotic event. Stochastic gradient descent is used to minimize the negative log

likelihood loss criterion, which is a measure that quantifies the classification error in the neural network.

The light design of the DNN, along with the nuclei sampling strategy, reduces the computational burden at test time as the cancerous regions in whole slide images can cover over 500 HPFs at 40× magnification. As such, an appropriate strategy combined with a suitable DNN architecture is necessary to meet the time constraints and requirements of a typical clinical workflow.

The DNN classifier was trained using image patches extracted from two open access datasets: 225 (from MITOS2012 [62]) and 516 mitotic events (from AMIDA2013 [66]). Non-mitotic candidate nuclei were extracted using an automatic nuclei detector. The performance of the DNN was evaluated with the AMIDA2013 training dataset under a 6-fold cross-validation setup and each fold was split at the patient level. The F_{score}, Precision and Recall measures [53] for the mitotic nuclei class were computed across the 6-folds. The average and the standard deviation for the F_{score}, precision and recall measures of the 6 folds were: $0{,}556 \pm 0{,}21$, $0{,}47 \pm 0{,}24$ and $0{,}78 \pm 0{,}11$, respectively. The mitotic classifier was also evaluated on the MITOS2012 testing set (using the AMIDA2013 and MITOS2012 training sets to train the DNN) obtaining a 0,78 F_{score}.

3.2.2. Support Vector Machine for BCa risk stratification

The DNN mitotic detector can be used on several HPFs from a WSI to assess the mitotic frequency in the respective cancerous regions. The obtained mitotic frequency information is then used to build a feature vector of 10 dimensions, each bin (dimension) representing the number of high power fields with [$k \in 0, 1..,8$, and $k \geq 9$] mitotic figures. Since the cancerous regions in each BCa case have a different number of HPFs, the feature vectors are normalized. Each feature vector is centered and scaled by a unitary standard deviation. The collected feature vectors were used to train a support vector machine with a linear kernel in order to distinguish BCa cases labeled either as low or high risk.

3.3. Experimental Design

3.3.1. Whole Slide Data Description

The 174 Whole Slide Images (WSI) are diagnosed with the Bloom Richardson (BR) grading system and ODX. The dataset was split into (a) High (24 WSI, $ODX > 30$), (b) Intermediate (55 cases and $18 \leq ODX \leq 30$) and (c) Low Risk categories (95 WSI and $ODX < 18$). Additionally, BR score defined another two risk subgroups: (d) 15 cases with both high ODX and high BR grade ($BR > 7$) and (e) 42 cases with both low ODX and low BR grade ($BR < 6$).

From the 174 WSI, a total of 20082 high power fields were extracted. Each high power field corresponds to an image of 2000×2000 pixels at $40\times$ magnification (a pixel is $\approx 0{,}0625\mu m^2$).

The performance of the DNN mitosis detector described in Subsection 'DNN for mitotic identification' was also evaluated in a subset of the images from these 174 WSI. A total of 40 HPFs were randomly selected from BCa cases that have either high ODX-high grade or low ODX-low grade. Mitoses in these images were annotated by an expert pathologist collaborator. Evaluation of the DNN performance yielded an overall 0,56 F-score measure in the aforementioned set of images. This measure is within the range previously reported ($0{,}556 \pm 0{,}21$) in Subsection 'DNN for mitotic identification'. A detailed presentation of the results is included in the Supplementary File B (See Appendix I).

3.3.2. Experiments and statistical analysis

The correlation of the automated mitosis with the ODX and BR groups was assessed in two different ways. First, a t-test evaluated the differences of the obtained mitotic counting for the different risk groups defined in Subsection 'Whole Slide Data Description'. Second, to evaluate the hypothesis that the automatically estimated mitotic count enables discrimination of the high/low ODX groups, a Support Vector Machine classifier was trained and used to classify WSI into high and low risk BCa.

A mean "mitotic count per ten high power fieldsïs computed for each WSI. The distribution is then compared for the two risk groups using a two tail t-test. The null hypothesis is that both distributions are sampled from the same normal distribution with equal mean. The t-test is applied using an equal variance for the different risk groups, specifically: (a) The high ODX group against the low ODX group, (b) The high ODX group against both the intermediate and low ODX group, (c) The high ODX-high grade (HH Group) against the low ODX-low grade score (LL group), (d) The high ODX-high grade (HH Group) against the rest of cases (HHc group) and (e) The low ODX-low grade (LL group) against the rest of cases (LLc group). The main hypothesis to test in each of the aforementioned experiments is that the BCa groups with higher ODX score and BR grading have a higher number of mitotic figures.

The discriminability of the number of mitotic figures automatically identified by the deep learning classifier was assessed in a BCa risk stratification task. The support vector machine classifier, described in Subsection 'Support Vector Machine for BCa risk stratification', was evaluated using a leave-one-out cross-validation scheme for each of the BCa cases (WSI) labeled either as high risk (BCa cases with high ODX score 24 cases) or low risk (BCa cases with low ODX score and low grade - 95 cases). The SVM classifier was trained with a regularization hyperparameter $C = 1$.

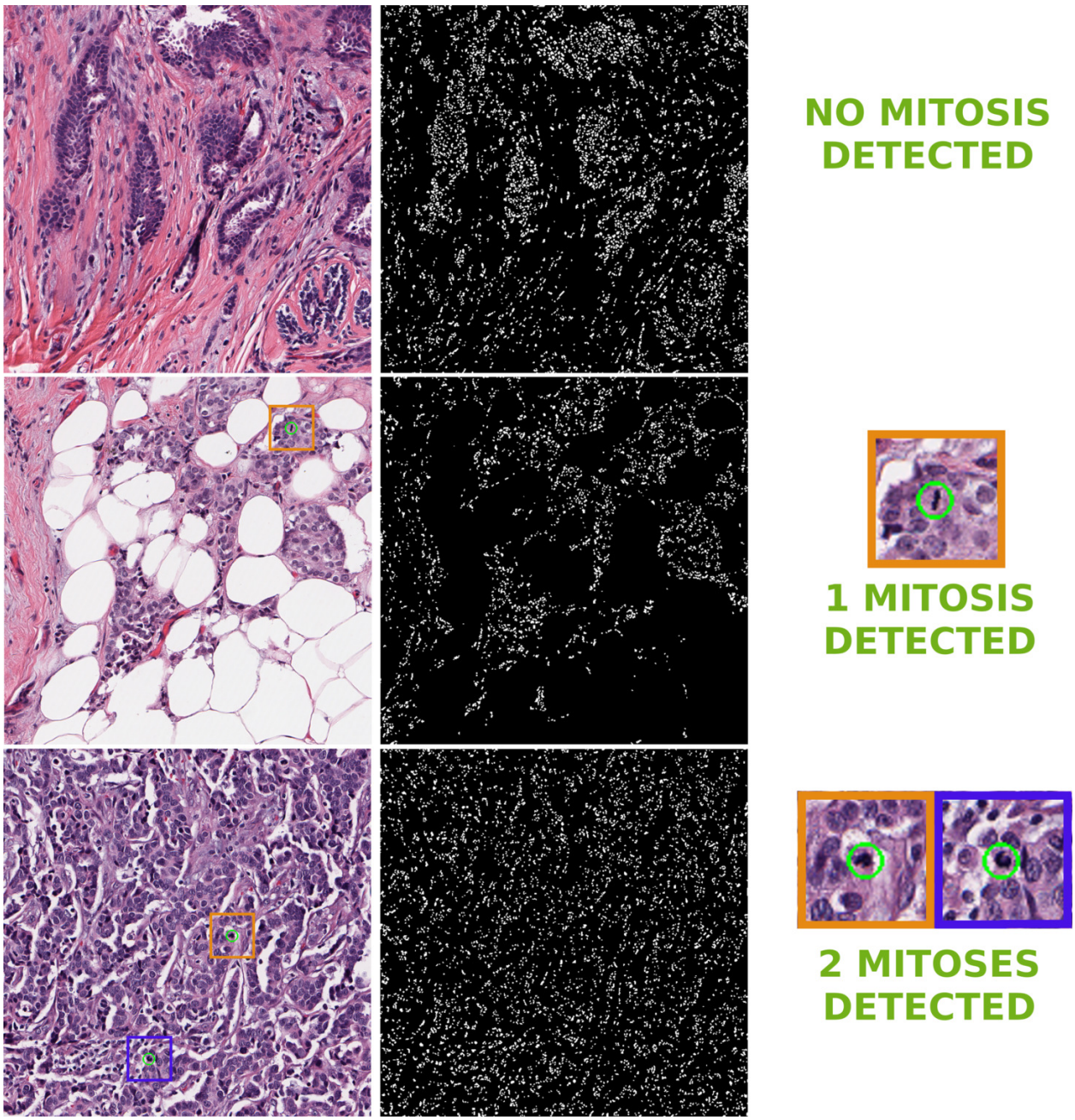

Figure 3-3.: Mitosis identification process in HPFs extracted from low ODX BCa (top row), intermediate ODX BCa (middle row) and high ODX BCa (bottom row). In the first column, results of the mitosis detection for the original HPF at $40X$ magnification is depicted. In the second column, the mask after the nuclei detection process is presented. The third column shows a close up for the DNN detection of the mitoses (green circles) depicted in the left most column. Mitosis are rare events: at the selected HPF size the low ODX Bca cases usually does not have mitosis. Meanwhile, intermediate and high ODX HPFs have a slightly larger mitosis count.

3.4. Experimental Results and Discussion

3.4.1. Statistical Analysis

The DNN classifier was applied to the 174 WSI previously described in Subsection 'Whole Slide Data Description'. Qualitative results for high, intermediate and low ODX cases can be seen in Figure **3-3**. Distributions of the average mitotic count for the high, intermediate and low ODX groups are depicted as violin plots in Figure **3-4** (see caption for violin plot description). Additionally, the intermediate group is paired either with the high or the low group to compare the resulting distributions.

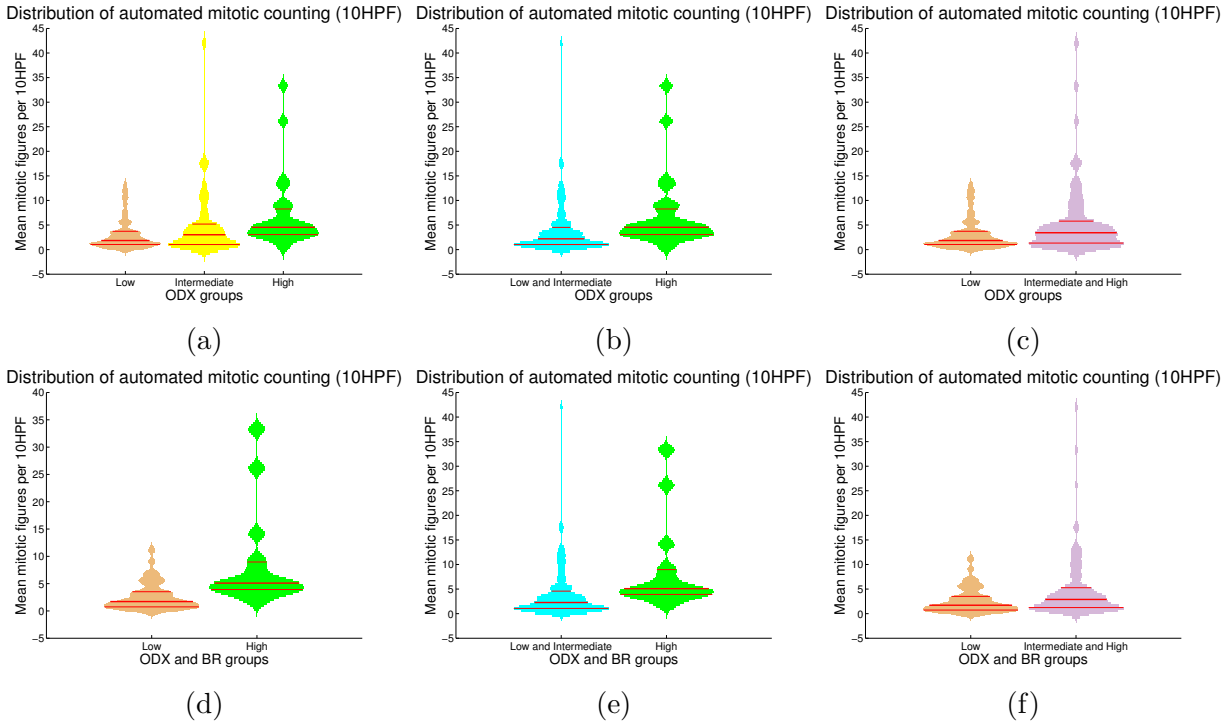

Figure 3-4.: Violin plots depicting the mean mitotic activity in 10 HPFs for the different risk ODX risk groups. The histogram associated with each violin plot is smoothed using a normal kernel. Red lines in the violin plot show the location of the lower quartile (q_1), the median and the upper quartile (q_3). Low (orange), intermediate (yellow) and high (green) ODX groups are shown in the top row (**a**). The mitotic distribution of the low and intermediate groups (cyan) against the high ODX group is presented in (**b**). The mitotic distribution of the low group against the intermediate and high ODX groups (purple) are presented in (**c**). The mitotic distribution for the groups with low ODX-low grade and high ODX-high grade are depicted in (**d**). The mitotic distribution for the high ODX-high grade against the rest of BCa cases, and low ODX-low grade against the rest of BCa cases are presented in (**e**) and (**f**) respectively.

The mean mitotic distribution in 10 HPF was computed for high ODX and BR scores (HH) and for low ODX and BR scores (LL), as illustrated in the bottom row of Figure **3-4**. The distributions of both groups was compared against the rest of BCa cases. The t-test results are presented in Table **3-1**.

Risk group comparison	**p-values (Equal variance)**
H vs L	$\mathbf{0{,}1 \times 10^{-3}}$
H vs L and I	$\mathbf{3{,}6 \times 10^{-3}}$
H and I vs L	$\mathbf{2{,}3 \times 10^{-3}}$
HH vs LL	$\mathbf{0{,}2 \times 10^{-3}}$
LL vs rest of Bca cases (LLc)	**0,032**
HH vs rest of Bca cases (HHc)	$\mathbf{0{,}6 \times 10^{-3}}$

Table 3-1.: Statistical evaluation of the DNN mitotic classifier in distinguishing different breast cancer ODX risk groups: High ODX risk group (H), Low ODX risk group (L) and Intermediate ODX risk group (I). Additionally, the Bloom Richardson grading scheme was also used to define the high ODX-high grade group (HH) and the low ODX-low grade group (LL). Note that statistically significant differences were observed for all the 6 comparative experiments performed.

Results in Figure **3-4** and Table **3-1** reveal that the average number of mitoses for the low ODX group is significantly different than those obtained for the high risk group. The low ODX group obtained a mean number of 3,1 mitoses per 10 HPFs while the high ODX group obtained a mean number of 7,2 mitoses per 10 HPFs ($p = 0{,}1 \times 10^{-3}$ with 95 % CI[$-6{,}2$, $-2{,}1$]). When grouping the intermediate ODX group with the low ODX group, the statistical test shows that the resultant mitoses distribution is different with respect to the high ODX group distribution, with a mean number of 3,8 mitoses per 10 HPFs for the low/intermediate ODX group ($p = 3{,}6 \times 10^{-3}$ with 95 % CI[$-5{,}8$, $-1{,}15$]). Besides, if the intermediate ODX group is combined with the high ODX group, the mixed distribution results different to the low ODX group distribution, with a mean number of 5,6 mitoses per 10 HPFs for the intermediate/high ODX group ($p = 2{,}3 \times 10^{-3}$ with 95 % CI[$-4{,}1$, $-0{,}9$]).

The analysis of the groups defined by both ODX score and BR grade also show a different number of mitotic figures. Cases with a low ODX-low grade yielded a mean number of 2,7 mitoses per 10 HPFs while those with high ODX-high grade showed a mean number of 8,8 mitoses per 10 HPFs ($p = 0{,}2 \times 10^{-3}$ with 95 %CI [$-9{,}2$,$-3{,}1$]). The low ODX-low grade cases continue to have a different number of mitoses when compared with all the remaining cases in the dataset (those that did not have both low ODX and low grade). In this case, the average number of mitoses per 10 HPFs for the 132 (LLc) cases was 4,74. The two distributions are closer but still significantly different ($p = 0{,}03$ with 95 %CI [$-3{,}98$, $-0{,}18$]). Finally, when comparing the high ODX-high grade cases with the remaining cases, a significant difference was found. For those cases that did not have both high ODX and high grade, the mean number of mitoses per 10HPFs was 3,8 ($p = 0{,}6 \times 10^{-3}$ with 95 %CI [$-7{,}8$,$-2{,}2$).

3.4.2. Support Vector Machine Classification Results

The distribution of mitoses per high power field was used to train a linear support vector machine classifier to discriminate the high/low ODX BCa cases, yielding an average accuracy of 83,19 %. Now, if one considers that the outcome of the leave one out evaluation in this case is binary, an error of the classification gives a null accuracy for that classifier. From the 119 different experiments, 20 resulted in a null accuracy. The confusion matrix, shown in Table **3-2**, was obtained by computing the aggregation of the confusion matrices over each fold along the leave-one-out experiment (119 experiments in this case). The results reveal that most of the WSIs with low ODX score are correctly classified as Low Risk. Leveraging a SVM classifier resulted in a mean increase of approximately 15 % in the observed accuracy (83,19 %) versus using a simple thresholding method (68,07 %). The details on how to compute the optimal threshold are described in the Supplementary File A (See Appendix H).

	Low ODX	***High ODX***
Classified as LR	83	8
Classified as HR	12	16

Table 3-2.: Confusion matrix for the risk stratification task using mitotic information. The BCa cases are classified either as low risk (LR) or high risk (HR).

3.5. Concluding Remarks

In this paper we rigorously investigated the problem of automated computation of the mitotic activity in WSI, a potential histologic image biomarker of the disease risk and aggressiveness in ER+ breast cancers. To evaluate whether the mitotic index was associated with the risk category determined by the Oncotype DX test, a DNN classifier was developed to identify mitoses. The mitotic count was then evaluated in terms of its ability to distinguish the low and high ODX risk categories as well as low and high grades. Likewise, the strategy was able to discriminate cases with different permutations of ODX risk and grade. On a cohort of 174 whole slide images, the automated mitotic count was significantly different for the low ODX group when compared against the high ODX risk group, thereby demonstrating that this quantification captures important prognostic information. This result also extends to the BCa cases group with both low ODX and low grade. On the other hand the mitotic detector failed to show differences between high and intermediate ODX groups.

The mitotic information extracted using the DNN achieves a sufficient level of precision and deals with patient tissue variability. The prognostic information is captured by the support vector machine classifier which properly distinguishes between BCa cases with both high ODX and low ODX score.

The lightweight DNN approach herein presented shows a similar performance to other DNN approaches, yet the particular sampling strategy herein integrated analyzes a HPF in around 8 seconds using a conventional GPU (Nvidia Quadro K4000). Other DNN approaches in the literature report a full HPF processing in about 8 minutes [16]. In a typical WSI, some of the cancerous regions may have more than 500 HPFs. Consequently, the lightweight DNN approach herein introduced is more appropriate for a diagnostic routine workflow.

The results of this study have significant clinical applicability even though the high and intermediate risk ODX groups cannot be separated via the DNN. Patients with low ODX have demonstrated low risk, and those patients are unlikely to derive a significant benefit from adjuvant chemotherapy. It is less clear that it is safe to omit chemotherapy in patients with intermediate and high risk scores, and oncologists routinely have to discuss the risks and benefits of chemotherapy in these circumstances.

Mitotic activity of a tumor, whether measured as number of mitoses per hight power field (as in traditional histopathologic grading), proliferation index as measured by Ki67 count by immunohistochemistry, or as measured by gene expression by predictive/prognostic assays such as ODX cannot be used alone to separate patients into low, intermediate or high risk categories. Mitotic activity is an important component of these analyses but it is not the lone determinant of the score. The fact that the mitotic activity figures from the DNN can alone separate low versus intermediate/high is quite remarkable and has direct clinical utility. It is possible that if this model were coupled with additional algorithms aimed at other pathologic features in the future, the risk category could be defined further.

Future work on improving the robustness of the mitotic detector to stain variation and including image information in the Z-axis (i.e pathologist frequently change the Z-axis of optical microscopes to determine if a confounding nuclei is mitotic) might improve the mitosis detector and allow for even more accurate estimation of mitotic rate. Also, future work will involve independent validation of these findings using a separate test cohort.

4. Conclusions

This thesis has investigated the correlation of automatically extracted breast cancer biomarkers in histological images and the cancer risk. It is notorious that grading systems in breast cancer (such as Bloom-Richardson system) use three histological criteria to assess the cancer risk: tubule formation, mitotic activity and nuclear pleomorphism. Hence, a hypothesis of this work is that automated strategies that quantify these criteria, or surrogate measures related to them, hold enough prognostic information to discriminate high/low risk breast cancers. This hypothesis was assessed by training a set of deep learning models to automatically identify mitoses and tubular nuclei in breast cancer cases. Firstly, the computation of a ratio between the number of tubule nuclei and the number of nuclei in the high power fields (Tubular Formation Index - TFI) was investigated as a potential automated image biomarker of the disease risk and aggressiveness. Secondly, the automated computation of the mitotic activity in WSI was also studied as a potential histologic image biomarker in ER+ breast cancers.

In the present investigation, the automated quantification of the TFI and the mitotic activity was correlated with the risk category determined by the Oncotype DX test. For doing so, a cohort of 174 breast cancer cases with their corresponding Oncotype DX recurrence score and their pathological grade were used. The TFI was significantly different for the BCa cases with low ODX-low histological grade and high ODX-high histological grade. When comparing the high ODX-high grade group with the remaining BCa cases, the TFI was still significantly lower. Likewise, the calculated tubule quantification measure was larger in the BCa cases with low ODX-low grade when compared to the remaining BCa cases. In the same cohort, the automated mitotic count was found to be significantly different for the low ODX group when compared against the high ODX risk group, demonstrating that automated quantification of mitoses captures important prognostic information. This result can also be extended to the BCa cases with both low ODX and low grade. The mitotic detector and the TFI nevertheless failed to show differences between high and intermediate ODX groups. The mitotic information extracted using the DNN achieves a sufficient level of precision and deals with patient tissue variability. The prognostic information is captured by the support vector machine classifier which properly distinguishes between BCa cases with high/low ODX score.

The computed TFI shows a slightly lower correlation with the ODX risk categories than the automated mitosis and other automated histological features such as nuclear architecture [6]. However, information of the different features is complementary and strategies to integrate

these features could potentially boost the accuracy of cancer risk assessment. Some previous work has successfully combined automated features (even extracted from differently stained samples from the same patient) and predicted the corresponding ODX risk category [5]. Therefore, fusion strategies to integrate information from different histological features (e.g. nuclear architecture, mitotic count, tubule density) is a future research path.

The usefulness of the automated quantification of tubule formation and mitotic activity in the clinical pathology workflows is still a pending issue to be demonstrated. These automated quantification tools aim to reduce the inter-reader variability and to quantify the cancer risk. These features could facilitate the streamlining and standardization of the clinical workflows. The presented quantification algorithms were used within cancerous regions previously selected by an expert pathologist. However, the automatic selection of diagnostically relevant regions in whole slide images is an open research topic on its own, and accurate delineation strategies have been recently introduced [19]. Future work should focus on improving the performance of tubule and mitosis detectors, validating the presented approach on larger test cohorts and incorporating automatic selection of region of interest.

The potential applicability of the results of this study to the clinical practice has been demonstrated even though high and intermediate ODX risk groups could not be completely separated. Patients with low ODX have demonstrated low risk, case in which the adjuvant chemotherapy could be unnecessary. It is less clear how safe it could be to omit chemotherapy in patients with intermediate and high risk scores, and oncologists routinely discuss about chemotherapy risks and benefits in these cases. The automated quantification of histological features (such as tubule formation or nuclear pleomorphism) might provide a clearer picture in such cases. The mitotic activity of a tumor, that can be estimated as the number of mitoses per high power field, or as the proliferation index by the Ki67 immuno-histochemistry count, or as the gene expression by predictive/prognostic assays such as ODX, cannot be used alone to determine if a breast cancer has low, intermediate or high risk of recurrence. Mitotic activity is an important feature but not the only histological criterion. The fact that the mitotic activity figures from the DNN can alone separate low versus intermediate/high is quite remarkable and has clinical utility. It is possible that if this model were coupled with additional algorithms specialized in other pathologic features, the risk category could be defined more accurately.

The deep learning algorithms used in this study presents several limitations, that once overcome could lead to more robust and accurate tubule/mitoses detections. The main disadvantage is that generalization from a small set of annotations to the whole populations is not a feasible task. The performance of detectors shows an important variability across H&E whole slide images with staining variations. Future work should focus on improving the robustness of the mitotic and tubule detectors in case of such stain variations

Bibliography

[1] Acs, Geza ; Kiluk, John ; Loftus, Loretta ; Laronga, Christine: Comparison of Oncotype DX and Mammostrat risk estimations and correlations with histologic tumor features in low-grade, estrogen receptor-positive invasive breast carcinomas. In: *Modern Pathology* 26 (2013), Nr. 11, S. 1451–1460

[2] Afework, A. ; Beynon, M. D. ; Bustamante, F. ; Demarzo, A. ; Ferreira, R. ; Miller, R. ; Silberman, M. ; Saltz, J. ; Sussman, A. ; Tsang, H.: Digital Dynamic Telepathology-The Virtual Microscope. In: *AMIA Annual Fall Symposium*, 1998

[3] Al-Kadi, Omar S.: Texture measures combination for improved meningioma classification of histopathological images. In: *Pattern Recognition* 43 (2010), Nr. 6, S. 2043 – 2053. – ISSN 0031–3203

[4] Allen, D. C. ; Allen, D. C. (Hrsg.) ; Cameron, R. I. (Hrsg.): *Histopathology specimens: clinical, pathological and laboratory aspects*. Springer Verlag, 2004

[5] Basavanhally, Ajay ; Feldman, Michael ; Shih, Natalie ; Mies, Carolyn ; Tomaszewski, John ; Ganesan, Shridar ; Madabhushi, Anant: Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: comparison to Oncotype DX. In: *J Pathol Inform* 2 (2011), Nr. Suppl S1:1

[6] Basavanhally, Ajay ; Xu, Jun ; Madabhushi, Anant ; Ganesan, Shridar: Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with oncotype DX assay. In: *I S Biomed Imaging*, 2009, S. 851–854

[7] Basavanhally, Ajay ; Yu, Elaine ; Xu, Jun ; Ganesan, Shridar ; Feldman, Michael ; Tomaszewski, John ; Madabhushi, Anant: Incorporating domain knowledge for tubule detection in breast histopathology using O'Callaghan neighborhoods. In: *SPIE Medical Imaging*, 2011, S. 796310

[8] Beck, Andrew H. ; Sangoi, Ankur R. ; Leung, Samuel ; Marinelli, Robert J. ; Nielsen, Torsten O. ; van de Vijver, Marc J. ; West, Robert B. ; van de Rijn, Matt ; Koller, Daphne: Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. In: *Science translational medicine* 3 (2011), Nr. 108, S. 108ra113–108ra113

[9] Belsare, A. D. ; Mushrif, M. M.: Histopathological Image Analysis Using Image Processing Techniques: An Overview. In: *Signal & Image Processing : An International Journal (SIPIJ)* vol. 3 No. 4 (2012), S. 23–36

[10] Bhargava, Rohit ; Madabhushi, Anant: A Review of Emerging Themes in Image Informatics and Molecular Analysis for Digital Pathology. In: *Annual Review of Biomedical Engineering* 18 (2016), Nr. 1

[11] Bueno, Gloria ; Gonzalez, Roberto ; Deniz, Oscar ; Gonzalez, Jesus ; Garcia-Rojo, Marcial: Colour model analysis for microscopic image processing. In: *Diagnostic Pathology* 3 (2008), Nr. Suppl 1, S. S18. – ISSN 1746–1596

[12] Bussolati, G.: Dissecting the pathologist's brain: mental processes that lead to pathological diagnoses. In: *Virchows Archiv* 448 (2006), S. 739–743

[13] Celis, R. ; Romo, D. ; Romero, E.: Blind colour separation of H&E stained histological images by linearly transforming the colour space. In: *Journal of Microscopy* 260 (2015), Nr. 3, S. 377–388. – ISSN 1365–2818

[14] Chang, Hang ; Fontenay, Gerald V. ; Han, Ju ; Cong, Ge ; Baehner, Frederick L. ; Gray, Joe W. ; Spellman, Paul T. ; Parvin, Bahram: Morphometic analysis of TCGA glioblastoma multiforme. In: *BMC bioinformatics* 12 (2011), Nr. 1, S. 484

[15] Chang, Hang ; Loss, Leandro A. ; Spellman, Paul T. ; Borowsky, Alexander ; Parvin, Bahram: Batch-invariant nuclear segmentation in whole mount histology sections. In: *I S Biomed Imaging*, 2012, S. 856–859

[16] Ciresan, Dan C. ; Giusti, Alessandro ; Gambardella, Luca M. ; Schmidhuber, JÃ$\frac{1}{4}$rgen: Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. In: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013* Bd. 8150. 2013, S. 411–418

[17] Crowley, R. S. ; Medvedeva, O.: An intelligent tutoring system for visual classification problem solving. In: *Artificial Intelligence In Medicine* 36 (2006), S. 85–117

[18] Crowley, R. S. ; Naus, G. J. ; III, J. S. ; Friedman, C. P.: Development of Visual Diagnostic Expertise in Pathology-An Information-processing Study. In: *Journal of the American Medical Informatics Association* 10 (2003), S. 39

[19] Cruz-Roa, Angel ; Basavanhally, Ajay ; González, Fabio ; Gilmore, Hannah ; Feldman, Michael ; Ganesan, Shridar ; Shih, Natalie ; Tomaszewski, John ; Madabhushi, Anant: Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: *SPIE Medical Imaging*, 2014, S. 904103

[20] Dalle, Jean-Romain ; Leow, Wee K. ; Racoceanu, D. ; Tutac, Adina E. ; Putti, T.C.: Automatic breast cancer grading of histopathological images. In: *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008. – ISSN 1557–170X, S. 3052–3055

[21] Dalton, Leslie W. ; Page, David L. ; Dupont, William D.: Histologic grading of breast carcinoma. A reproducibility study. In: *Cancer* 73 (1994), Nr. 11, S. 2765–2770

[22] Demir, C. ; Yener, B.: Automated cancer diagnosis based on histopathological images: A systematic survey. / Computer Science Department at Rensselaer Polytechnic Institute. 2005 (Technical Report TR-05-09.). – Forschungsbericht

[23] Doyle, S. ; Feldman, M. ; Tomaszewski, J. ; Madabhushi, A.: A Boosted Bayesian Multiresolution Classifier for Prostate Cancer Detection From Digitized Needle Biopsies. In: *Biomedical Engineering, IEEE Transactions on* 59 (2012), Nr. 5, S. 1205–1218. – ISSN 0018–9294

[24] Doyle, S. ; Hwang, M. ; Shah, K. ; Madabhushi, A. ; Feldman, M. ; Tomaszeweski, J.: Automated Grading of Prostate Cancer Using Architectural and Textural Image Features. In: *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, 2007, S. 1284–1287

[25] Drake, Thomas A. ; Braun, Jonathan ; Marchevsky, Alberto ; Kohane, Isaac S. ; Fletcher, Christopher ; Chueh, Henry ; Beckwith, Bruce ; Berkowicz, David ; Kuo, Frank ; Zeng, Qing T. ; Balis, Ulysses ; Holzbach, Ana ; McMurry, Andrew ; Gee, Connie E. ; McDonald, Clement J. ; Schadow, Gunther ; Davis, Mary ; Hattab, Eyas M. ; Blevins, Lonnie ; Hook, John ; Becich, Michael ; Crowley, Rebecca S. ; Taube, Sheila E. ; Berman, Jules: A system for sharing routine surgical pathology specimens across institutions: the Shared Pathology Informatics Network. In: *Human Pathology* 38 (2007), August, S. 1212–1225

[26] Elston, C.W. ; Ellis, I.O.: 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. In: *Histopathology* 19 (1991), Nr. 5, S. 403–410. – ISSN 1365–2559

[27] Fan, Fang ; Thomas, PatriciaA.: Tumors of the Breast. In: Damjanov, Ivan (Hrsg.) ; Fan, Fang (Hrsg.): *Cancer Grading Manual*. Springer New York, 2007. – ISBN 978–0–387–33750–0, S. 75–81

[28] Ferreira, R. ; Moon, B. ; Humphries, J. ; Sussman, A. ; Saltz, J. ; Miller, R. ; Demarzo, A.: The Virtual Microscope. In: *American Medical Informatics Association, 1997 Annual Fall Symposium*, 1997, S. 449–453

[29] Flanagan, Melina B. ; Dabbs, David J. ; Brufsky, Adam M. ; Beriwal, Sushil ; Bhargava, Rohit: Histopathologic variables predict Oncotype DX recurrence score. In: *Modern Pathology* 21 (2008), Nr. 10, S. 1255–1261

[30] Fontelo, P. ; DiNino, E. ; Krista, J. ; Khan, A. ; Ackerman, M.: Virtual Microscopy: Potential Applications in Medical Education and Telemedicine in Countries with Developing Economies. In: *Proceedings of the 38th Hawaii International Conference on System Sciences*, 2005

[31] Ghaznavi, Farzad ; Evans, Andrew ; Madabhushi, Anant ; Feldman, Michael: Digital imaging in pathology: whole-slide imaging and beyond. In: *Annual Review of Pathology: Mechanisms of Disease* 8 (2013), S. 331–359

[32] Gu, J. ; Ogilvie, R. W.: *Virtual Microscopy and Virtual Slides in Teaching, Diagnosis, and Research*. CRC Press, 2005

[33] Gurcan, Metin N. ; Boucheron, Laura E. ; Can, Ali ; Madabhushi, Anant ; Rajpoot, Nasir M. ; ; Yener, Bulent: Histopathological Image Analysis: A Review. In: *IEEE Reviews in Biomedical Engineering* vol. 2 (2009), S. 147–171

[34] Gurcan, M.N. ; Pan, T. ; Shimada, H. ; Saltz, J.: Image Analysis for Neuroblastoma Classification: Segmentation of Cell Nuclei. In: *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, 2006. – ISSN 1557–170X, S. 4844–4847

[35] Ioffe, Sergey ; Szegedy, Christian: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *arXiv preprint arXiv:1502.03167* (2015)

[36] Iregui, Marcela ; Gomez, Francisco ; Romero, Eduardo: Virtual Microscopy in Medical Image: a Survey. In: *Modern Research and Educational Topics in Microscopy* 2 (2007), S. 996–1006

[37] Janowczyk, Andrew ; Madabhushi, Anant: Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. In: *Journal of Pathology Informatics* 7 (2016)

[38] Khoury, Thaer ; Huang, Xiao ; Chen, Xiwei ; Wang, Dan ; Liu, Song ; Opyrchal, Mateusz: Comprehensive Histologic Scoring to Maximize the Predictability of Pathology-generated Equation of Breast Cancer Oncotype DX Recurrence Score. In: *Appl Immunohisto M M* (2016)

[39] Klein, Molly E. ; Dabbs, David J. ; Shuai, Yongli ; Brufsky, Adam M. ; Jankowitz, Rachel ; Puhalla, Shannon L. ; Bhargava, Rohit: Prediction of the Oncotype DX recurrence score: use of pathology-generated equations derived by linear regression analysis. In: *Modern Pathology* 26 (2013), S. 658–664

[40] Krishnan, M.M.R. ; Shah, P. ; Ghosh, M. ; Pal, M. ; Chakraborty, C. ; Paul, R.R. ; Chatterjee, J. ; Ray, A.K.: Automated characterization of sub-epithelial connective tissue cells of normal oral mucosa: Bayesian approach. In: *Students' Technology Symposium (TechSym), 2010 IEEE*, 2010, S. 44–48

[41] Krizhevsky, Alex ; Sutskever, Ilya ; Hinton, Geoffrey E.: Imagenet classification with deep convolutional neural networks. In: *Adv Neur In*, 2012, S. 1097–1105

[42] Krupinski, E. A. ; Tillack, A. A. ; Richter, L. ; Henderson, J. T. ; Bhattacharyya, A. K. ; Scott, K. M. ; Graham, A. R. ; Descour, M. R. ; Davis, J. R. ; Weinstein, R. S.: Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. In: *Human pathology* 37 (2006), Nr. 12, S. 1543–1556

[43] Larsen, Anders Boesen L. ; Vestergaard, Jacob S. ; Larsen, Rasmus: HEp-2 cell classification using shape index histograms with donut-shaped spatial pooling. In: *IEEE T Med Imaging* 33 (2014), Nr. 7, S. 1573–1580

[44] Lee, George ; Ali, Sahirzeeshan ; Veltri, Robert ; Epstein, Jonathan I. ; Christudass, Christhunesa ; Madabhushi, Anant: Cell orientation entropy (COrE): Predicting biochemical recurrence from prostate cancer tissue microarrays. In: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*, Springer, 2013, S. 396–403

[45] Lee, George ; Sparks, Rachel ; Ali, Sahirzeeshan ; Shih, Natalie N. ; Feldman, Michael D. ; Spangler, Elaine ; Rebbeck, Timothy ; Tomaszewski, John E. ; Madabhushi, Anant: Co-Occurring Gland Angularity in Localized Subgraphs: Predicting Biochemical Recurrence in Intermediate-Risk Prostate Cancer Patients. In: *PloS one* 9 (2014), Nr. 5, S. 1–14

[46] Lewis Jr, James S. ; Ali, Sahirzeeshan ; Luo, Jingqin ; Thorstad, Wade L. ; Madabhushi, Anant: A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. In: *Am J Surg Pathol* 38 (2014), Nr. 1, S. 128–137

[47] Madabhushi, Anant ; Lee, George: Image Analysis and Machine Learning in Digital Pathology: Challenges and Opportunities. In: *Medical Image Analysis* (2016)

[48] Maqlin, P ; Thamburaj, Robinson ; Mammen, Joy J. ; Nagar, Atulya K.: Automatic detection of tubules in breast histopathological images. In: *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BICTA 2012)* Springer, 2013, S. 311–321

[49] Meyer, John S. ; Alvarez, Consuelo ; Milikowski, Clara ; Olson, Neal ; Russo, Irma ; Russo, Jose ; Glass, Andrew ; Zehnbauer, Barbara A. ; Lister, Karen ; Parwaresch, Reza: Breast carcinoma malignancy grading by Bloom–Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. In: *Modern pathology* 18 (2005), Nr. 8, S. 1067–1078

[50] Moncayo, Ricardo ; Romo-Bucheli, David ; Romero, Eduardo: A Grading Strategy for Nuclear Pleomorphism in Histopathological Breast Cancer Images Using a Bag of Features (BOF). In: *Iberoamerican Congress on Pattern Recognition* Springer, 2015, S. 75–82

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

[52] Otsu, Nobuyuki: A Threshold Selection Method from Gray-Level Histograms. In: *IEEE T Syst Man Cyb* 9 (1979), Jan, Nr. 1, S. 62–66

[53] Powers, David M.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. In: *J Mach Learn Tech* (2011)

[54] Roa-Peña, Lucia ; Gómez, Francisco ; Romero, Eduardo: An experimental study of pathologist's navigation patterns in virtual microscopy. In: *Diagnostic pathology* 5 (2010), Nr. 1, S. 71

[55] Romo, David ; GarcÃa-Arteaga, Juan D. ; ArbelÃ¡ez, Pablo ; Romero, Eduardo: A discriminant multi-scale histopathology descriptor using dictionary learning, 2014, S. 90410Q–90410Q–6

[56] Romo, David E. ; Tarquino, Jonathan ; GarcÃa-Arteaga, Juan D. ; Romero, Eduardo: Virtual slide mosaicing using feature descriptors and a registration consistency measure, 2013, S. 89220Q–89220Q–8

[57] Romo-Bucheli, D. ; Moncayo, R. ; Cruz-Roa, A. ; Romero, E.: Identifying histological concepts on basal cell carcinoma images using nuclei based sampling and multi-scale descriptors. In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015. – ISSN 1945–7928, S. 1008–1011

[58] Romo-Bucheli, David ; Corredor, GermÃ¡n ; GarcÃa-Arteaga, Juan D. ; Arias, Viviana ; Romero, Eduardo: Nuclei graph local features for basal cell carcinoma classification in whole slide images, 2017, S. 101600Q–101600Q–9

[59] Romo-Bucheli, David ; Janowczyk, Andrew ; Gilmore, Hannah ; Romero, Eduardo ; Madabhushi, Anant: Automated Tubule Nuclei Quantification and Correlation with Oncotype DX risk categories in ER+ Breast Cancer Whole Slide Images. In: *Scientific Reports* 6 (2016), Sep, S. 32706 EP –. – Article

[60] Romo-Bucheli, David ; Janowczyk, Andrew ; Gilmore, Hannah ; Romero, Eduardo ; Madabhushi, Anant: A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. In: *Cytometry Part A* (2017), S. n/a–n/a. – ISSN 1552–4930

[61] Romo-Bucheli, David ; Janowczyk, Andrew ; Romero, Eduardo ; Gilmore, Hannah ; Madabhushi, Anant: Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images, 2016, S. 979106–979106–11

[62] Roux, Ludovic ; Racoceanu, Daniel ; Loménie, Nicolas ; Kulikova, Maria ; Irshad, Humayun ; Klossa, Jacques ; Capron, Frederique ; Genestie, Catherine ; Le Naour, Gilles ; Gurcan, Metin N.: Mitosis detection in breast cancer histological images An ICPR 2012 contest. In: *J Pathol Inform* 4 (2013)

[63] Sertel, O. ; Catalyurek, U.V. ; Shimada, H. ; Guican, M.N.: Computer-aided Prognosis of Neuroblastoma: Detection of mitosis and karyorrhexis cells in digitized histological images. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, 2009. – ISSN 1557–170X, S. 1433–1436

[64] Tambasco, Mauro ; Eliasziw, Misha ; Magliocco, Anthony M.: Morphologic complexity of epithelial architecture for predicting invasive breast cancer survival. In: *Journal of translational medicine* 8 (2010), Nr. 1, S. 1–10

[65] Thaker, Nikhil G. ; Hoffman, Karen E. ; Stauder, Michael C. ; Shaitelman, Simona F. ; Strom, Eric A. ; Tereffe, Welela ; Smith, Benjamin D. ; Perkins, George H. ; Huo, Lei ; Munsell, Mark F. [u. a.]: The 21-gene recurrence score complements IBTR! Estimates in early-stage, hormone receptor-positive, HER2-normal, lymph node-negative breast cancer. In: *SpringerPlus* 4 (2015), Nr. 1, S. 36

[66] Veta, Mitko ; van Diest, Paul J. ; Willems, Stefan M. ; Wang, Haibo ; Madabhushi, Anant ; Cruz-Roa, Angel ; Gonzalez, Fabio ; Larsen, Anders B. ; Vestergaard, Jacob S. ; Dahl, Anders B. [u. a.]: Assessment of algorithms for mitosis detection in breast cancer histopathology images. In: *Medical image analysis* 20 (2015), Nr. 1, S. 237–248

[67] Veta, Mitko ; Pluim, Josien P. ; van Diest, Paul J. ; Viergever, Max [u. a.]: Breast cancer histopathology image analysis: A review. In: *IEEE T Bio-med Eng* 61 (2014), Nr. 5, S. 1400–1411

[68] Waheed, S. ; Moffitt, R.A. ; Chaudry, Q. ; Young, A.N. ; Wang, M.D.: Computer Aided Histopathological Classification of Cancer Subtypes. In: *Bioinformatics and Bioengineering, 2007. BIBE 2007. Proceedings of the 7th IEEE International Conference on*, 2007, S. 503–508

[69] Wang, Haibo ; Cruz-Roa, Angel ; Basavanhally, Ajay ; Gilmore, Hannah ; Shih, Natalie ; Feldman, Mike ; Tomaszewski, John ; Gonzalez, Fabio ; Madabhushi, Anant: Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. In: *J Med Imag* 1 (2014), Nr. 3, S. 034003