



UNIVERSIDAD NACIONAL DE COLOMBIA

# **Data-driven Representation Learning from Histopathology Image Databases to Support Digital Pathology Analysis**

**Angel Alfonso Cruz-Roa**

Universidad Nacional de Colombia  
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial  
Bogotá D.C., Colombia  
2015

# **Data-driven Representation Learning from Histopathology Image Databases to Support Digital Pathology Analysis**

**Angel Alfonso Cruz-Roa**

In fulfillment of the requirements for the degree of:  
**Doctor en Ingeniería - Ingeniería de Sistemas y Computación**

Advisor:  
Fabio Augusto González Osorio, Ph.D.

Research Field:  
Image Representation - Feature and Representation Learning - Medical Image Analysis  
Research Group:  
Machine learning, perception and discovery Lab - MindLab  
BioIngenium Research Group

Universidad Nacional de Colombia  
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial  
Bogotá D.C., Colombia  
2015

A los cuatro pilares de mi vida,  
A mi esposa Marcela  
A mis padres Alfonso y Nubia  
A mi hermano Andrés.  
Sin su incondicional apoyo y comprensión...  
nada de esto habría sido posible

*"If I have seen further it is by standing on the shoulders of giants".*  
*"Si he logrado ver más lejos, ha sido porque he subido a hombros de gigantes".*  
— **Sir Isaac Newton**

# Acknowledgements

I want to express my gratitude to my advisor, professor Fabio González, for his support, guidance, and patience. During this process, I learned a lot of valuable lessons, abilities and skills for both academic and personal life which I expect to reproduce as best as possible. The most important lessons have been the commitment as mentor and encouragement to show how can we go beyond our limits by exploiting all our potential.

My thanks also are to professor Eduardo Romero, his vision, support and advice were always present during all this time give me an additional point of view for my formation. Special thanks to professor Anant Madabhushi whom receive me twice during my internships in Center of Computational Imaging and Personalized Diagnostics (CCIPD) at Case Western Reserve University, his perspective and research dynamics complemented my learning and research process. I learn a lot from them about the importance to tackle biomedical problems that could impact the society and how it is possible to connect academy and industry by proposing innovative technological solutions.

Thanks to all partners and friends from former Bioingenium Research Group, and now Cim@Lab and MindLab groups. My sincere thanks to Juan Caicedo, Jorge Camargo, John Arévalo, Fabio Martínez, Andrea Rueda, Gloria Díaz, Jorge Vanegas, David Romo, Ricardo Gutiérrez, Raúl Ramos, Carlos Vargas, Francisco Gómez, Fabián Narváez, Hugo Franco and Juan García. Thanks also to people from CCIPD at Case Western Reserve University during my both internships in United States. I shared great moments with Mirabela Rusu, Ajay Basavanhally, Haibo Wang, Madhi Oroji, Sahir Ali, Ahmad Algohary and George Lee. Special thanks also for all pathologists and physicians whom teach me a lot about the challenging pathology workflow during this process. My thanks go to Lucia Peña, Viviana Arias, Hannah Gilmore, Michael Feldman and Shridar Ganesan.

I also want to thank to the institutions that have allowed and supported this thesis. This work have been funded by a doctoral fellowship grant 528/2011 and projects 1101-521-28803, 1101-521-28767 and 1225-569-34920 from the Administrative Department of Science, Technology and Innovation of Colombia (Colciencias). Also thanks to Uni-

versidad Nacional de Colombia, their professors, infrastructure and funding support through different dependences, such as, Doctoral Program on Systems and Computing Engineering, *Coordinación curricular de posgrados de Ing. Sistemas e Industrial*, *Vicedecanatura de Investigación y Extensión* from School of Engineering, *Dirección de Bienestar de Sede Bogotá*, *Dirección de Investigación Sede Bogotá* by funding for doctoral thesis project and international mobility. Finally, also thanks to Case Western Reserve University by funding provided in the two research visits to Cleveland, OH (USA).

Finally, very special thanks to my parents, whom were the very first persons that I ask about to start my PhD and they engage me to do it. They always support me and I will do not have enough time to get back everything that they did for me. My infinite grateful to my lovely wife, who have been my main support in all moments including the most challenging ones, she believes more in me than myself. Thanks to my friend and brother, he always has been present to share nice and fun moments. Very special thanks to all my family, and who they are no longer with us. Many times they have to be patience when I had to sacrifice personal time sharing with them for pursuit this goal.

# Abstract

Cancer research is a major public health priority in the world due to its high incidence, diversity and mortality. Despite great advances in this area during recent decades, the high incidence and lack of specialists have proven that one of the major challenges is to achieve early diagnosis. Improved early diagnosis, especially in developing countries, plays a crucial role in timely treatment and patient survival. Recent advances in scanner technology for the digitization of pathology slides and the growth of global initiatives to build databases for cancer research have enabled the emergence of *digital pathology* as a new approach to support pathology workflows. This has led to the development of many computational methods for automatic histopathology image analysis, which in turn has raised new computational challenges due to the high visual variability of histopathology slides, the difficulty in assessing the effectiveness of methods (considering the lack of annotated data from different pathologists and institutions), and the need of interpretable, efficient and feasible methods for practical use. On the other hand, machine learning techniques have focused on exploiting large databases to automatically extract and induce information and knowledge, in the form of patterns and rules, that allow to connect low-level content with its high-level meaning. Several approaches have emerged as opposed to traditional schemes based on handcrafted features for data representation, which nowadays are known as *representation learning*. The objective of this thesis is the exploration, development and validation of precise, interpretable and efficient computational machine learning methods for automatic representation learning from histopathology image databases to support diagnosis tasks of different types of cancer. The validation of the proposed methods during the thesis development allowed to corroborate their capability in several histopathology image analysis tasks of different types of cancer. These methods achieve good results in terms of accuracy, robustness, reproducibility, interpretability and feasibility suggesting their potential practical application towards translational and personalized medicine.

**Keywords:** Digital pathology, histopathology image analysis, representation learning, deep learning, whole slide images.

# Resumen

La investigación en cáncer es una de las principales prioridades de salud pública en el mundo debido a su alta incidencia, diversidad y mortalidad. A pesar de los grandes avances en el área en las últimas décadas, la alta incidencia y la falta de especialistas ha llevado a que una de las principales problemáticas sea lograr su detección temprana, en especial en países en vías de desarrollo, como quiera a que de ello depende las posibilidades de un tratamiento oportuno y las oportunidades de supervivencia de los pacientes. Los recientes avances en tecnología de escáneres para digitalización de láminas de patología y el crecimiento de iniciativas mundiales para la construcción de bases de datos para la investigación en cáncer, han permitido el surgimiento de la *patología digital* como un nuevo enfoque para soportar los flujos de trabajo en patología. Esto ha llevado al desarrollo de una gran variedad de métodos computacionales para el análisis automático de imágenes de histopatología, lo cual ha planteado nuevos desafíos computacionales debido a la alta variabilidad visual de las láminas de histopatología; la dificultad para evaluar la efectividad de los métodos por la falta de datos de diferentes instituciones que cuenten con anotaciones por parte de los patólogos, y la necesidad de métodos interpretables, eficientes y factibles para su uso práctico. Por otro lado, el aprendizaje de máquina se ha enfocado en explotar las grandes bases de datos para extraer e inducir de manera automática información y conocimiento, en forma de patrones y reglas, que permita conectar el contenido de bajo nivel con su significado. Diferentes técnicas han surgido en contraposición a los esquemas tradicionales basados en diseño manual de la representación de los datos, en lo que se conoce como *aprendizaje de la representación*. El propósito de esta tesis fue la exploración, desarrollo y validación de métodos computacionales de aprendizaje de máquina precisos, interpretables y eficientes a partir de bases de datos de imágenes de histopatología para el aprendizaje automático de la representación en tareas de apoyo al diagnóstico de distintos tipos de cáncer. La validación de los distintos métodos propuestos durante el desarrollo de la tesis permitieron corroborar la capacidad de cada uno de ellos en distintivas tareas de análisis de imágenes de histopatología, en diferentes tipos de cáncer, con buenos resultados en términos de exactitud, robustez, reproducibilidad, interpretabilidad y factibilidad, lo cual sugiere su potencial aplicación práctica hacia la medicina traslacional y personalizada.

**Palabras clave:** Patología digital, análisis de imágenes de histopatología, aprendizaje de la representación, aprendizaje profundo, láminas virtuales de patología.

# Contents

<b>List of Figures</b>	<b>xi</b>
------------------------	-----------

<b>List of Tables</b>	<b>xvii</b>
-----------------------	-------------

<b>1 Introduction</b>	<b>2</b>
1.1 Motivation . . . . .	2
1.2 Research Problem . . . . .	6
1.2.1 Research Question . . . . .	8
1.3 Challenges . . . . .	10
1.4 Contributions and Academic Products . . . . .	10
1.4.1 Semantic Part-based image Representations . . . . .	10
1.4.2 Feature and Representation Learning . . . . .	12
1.4.3 Accurate and high-throughput whole-slide histopathology image analysis . . . . .	14
1.4.4 Other contributions . . . . .	14
1.5 Organization of this thesis . . . . .	16

<b>I Semantic Part-based image Representations</b>	<b>19</b>
--	-----------

<b>2 Automatic Annotation of Histopathological Images with a Latent Topic Model</b>	<b>20</b>
2.1 Histopathology Image Annotation in Basal-cell Carcinoma . . . . .	20
2.1.1 Introduction . . . . .	21
2.1.2 Methodology . . . . .	23
2.1.3 Experimental design . . . . .	29
2.1.4 Results and discussion . . . . .	32
2.1.5 Conclusions . . . . .	35
2.2 Anaplastic Tumor Differentiation and Visual Interpretation in Medulloblastoma . . . . .	36
2.2.1 Introduction . . . . .	36



2.2.2	Methodology . . . . .	38
2.2.3	Experimental design . . . . .	41
2.2.4	Results and discussion . . . . .	42
2.2.5	Conclusions . . . . .	45
<b>II</b>	<b>Feature and Representation Learning</b>	<b>46</b>
<b>3</b>	<b>Unsupervised histopathology image representation learning</b>	<b>47</b>
3.1	Tumor Classification of Basal Cell Carcinoma . . . . .	47
3.1.1	Introduction . . . . .	48
3.1.2	Methodology . . . . .	50
3.1.3	Experimental design . . . . .	53
3.1.4	Results and discussion . . . . .	54
3.1.5	Conclusions . . . . .	57
3.2	Tumor Differentiation of Medulloblastoma Brain Cancer . . . . .	57
3.2.1	Introduction . . . . .	58
3.2.2	Methodology . . . . .	59
3.2.3	Experimental design . . . . .	62
3.2.4	Results and discussion . . . . .	63
3.2.5	Conclusions . . . . .	63
<b>4</b>	<b>Supervised histopathology image representation learning</b>	<b>65</b>
4.1	Mitosis detection in Breast Cancer . . . . .	65
4.1.1	Introduction . . . . .	66
4.1.2	Methodology . . . . .	69
4.1.3	Experimental design . . . . .	73
4.1.4	Results and discussion . . . . .	74
4.1.5	Conclusions . . . . .	77
4.2	Invasive BCa detection in Breast Cancer . . . . .	78
4.2.1	Introduction . . . . .	80
4.2.2	Methodology . . . . .	81
4.2.3	Experimental design . . . . .	83
4.2.4	Results and discussion . . . . .	85
4.2.5	Conclusions . . . . .	96
<b>III</b>	<b>Accurate and high-throughput whole-slide histopathology im- age analysis</b>	<b>97</b>

---

<b>5</b>	<b>Scaling up convolutional networks through high-throughput adaptive sampling: Applications in Digital Pathology</b>	<b>98</b>
5.1	High-throughput whole-slide image analysis in Invasive Breast Cancer .	98
5.1.1	Introduction . . . . .	99
5.1.2	Previous and related work . . . . .	101
5.1.3	Methodology . . . . .	103
5.1.4	Experimental design . . . . .	110
5.1.5	Results and discussion . . . . .	113
5.1.6	Conclusions . . . . .	118
<b>6</b>	<b>Conclusions and Perspectives</b>	<b>123</b>
6.1	Perspectives . . . . .	124

# List of Figures

<b>1-1</b>	Common digital pathology workflow. . . . .	5
<b>1-2</b>	Histotechnique procedure and digitalization process. From top left to bottom right: 1) Small cuts are made from the tissue 2-3) The smaller cuts are put into cassettes and (after processing) embedded in paraffin 4) Thin sections are made from the paraffin blocks with a microtome 5) The sections are put onto glass slides for staining 6) H&E stained slide which is 7) digitized with a pathology slide scanner to obtain 8) a whole-slide image. (Adapted from: <a href="http://amida13.isi.uu.nl/?q=node/2">http://amida13.isi.uu.nl/?q=node/2</a> )	7
<b>1-3</b>	Visual variability of several normal and pathological patterns in histopathology images in different cancer types (skin, breast and brain). . .	9
<b>2-1</b>	Example of histopathological images globally annotated with multiple annotations (multi-labeled images). These images correspond to the test data set used in this work and they have a resolution of $1024 \times 768$ pixels. Histopathological annotations of morphological and architectural features such as epidermis, collagen and hair follicles appear in different images illustrating the high visual variability for the same annotation. .	22
<b>2-2</b>	Overview of proposed method for automatic annotation of histopathological images based on non-negative matrix factorization. . . . .	24
<b>2-3</b>	Bag of feature setup used for representing histopathology images. In this work the local features extraction is performed using regular grid extraction and each patch of $8 \times 8$ pixels is represented by the first coefficients of a discrete cosine transform applied to each color component (RGB) independently, the visual codebook is built using $k$ -means with $k = 700$ , and finally each image is represented by a histogram of 700 bins normalized with $L_1$ norm. . . . .	26
<b>2-4</b>	Examples of training images with the corresponding histopathological annotations. These images have a resolution of $300 \times 300$ pixels and exhibit only one annotation per image. . . . .	31
<b>2-5</b>	Performance evaluation on training mono-label images by each number of dimensions in the latent space. . . . .	34

<b>2-6</b>	Example of an image from the test data set automatically annotated by the proposed method. The original multi-label image (a) is showed with the salient maps of the patches inside the image according with each one of the ten histopathological annotations: <i>collagen</i> (b), <i>sebaceous glands</i> (c), <i>hair follicles</i> (d), <i>inflammatory infiltration</i> (e), <i>eccrine glands</i> (f), <i>epidermis</i> (g), <i>nodular basal cell carcinoma</i> (h), <i>morpheiform basal cell carcinoma</i> (i), <i>micro-nodular basal cell carcinoma</i> (j), <i>cystic basal cell carcinoma</i> (k)), on the top of each salient image is the real membership of the class (v), the conditional probability estimated by the proposed method (p) and the final concept binarization value (b). . . . .	35
<b>2-7</b>	Overall scheme for the visual latent semantic analysis for automatic classification and interpretation method. . . . .	39
<b>2-8</b>	Dictionary size vs average accuracy for both local features (block and haar). . . . .	42
<b>2-9</b>	Visual semantic maps for test images from a cross validation trial. Images in the first column were correctly classified (A-anaplastic and NA-non-anaplastic). See Subsection 2.2.4 for description. . . . .	44
<b>3-1</b>	Example of BCC histopathology images (both cancer and non-cancer) stained with H&E at 10X. . . . .	49
<b>3-2</b>	Convolutional auto-encoder neural network architecture for histopathology image representation learning, automatic cancer detection and visually interpretable prediction results analogous to a <i>digital stain</i> identifying image regions that are most relevant for diagnostic decisions. . .	51
<b>3-3</b>	Comparison of learned features (a.k.a. dictionaries or basis) by autoencoders from: a) BCC (histopathology), b) HistologyDS (healthy tissues) and c) STL-10 (natural scenes) datasets, and d) DCT basis. . . . .	55
<b>3-4</b>	Overview of representation learning framework for medulloblastoma differentiation between anaplastic and non-anaplastic tumors by comparing unsupervised and supervised feature learning approaches. . . . .	60
<b>4-1</b>	An illustration of the visual similarity between true mitotic processes and confounding false positives. (a)-(c): true mitoses. (d)-(f): confounding non-mitotic figures. . . . .	67

- 4-2** Workflow of our methodology. Blue-ratio thresholding [23] is first applied to segment mitosis candidates. On each segmented blob, handcrafted features are extracted and classified via a Random Forests classifier. Meanwhile, on each segmented  $80 \times 80$  patch, convolutional neural networks (CNN) [102] are trained with a fully connected regression model as part of the classification layer. For those candidates that are difficult to classify (ambiguous result from the CNN), we train a second-stage Random Forests classifier on the basis of combining CNN-derived and handcrafted features. Final decision is obtained via a consensus of the predictions of the three classifiers. . . . . 69
- 4-3** Example of blue-ratio segmentation. (a) is the original HPF slice while (b) is the segmentation mask. Note that a majority of the objects identified via this approach in (b) are indeed mitotic figures. . . . . 70
- 4-4** Architecture of the CNN model. The CNN architecture comprises 3 layers: two consecutive convolutional-pooling layers and a fully-connected classification layer. The two convolutional-pooling layers use the same fixed  $8 \times 8$  convolutional kernel and  $2 \times 2$  pooling kernel, but have 64 and 128 neurons, respectively. The last layer has 256 neurons, which are all connected to the final two neurons for mitosis/non-mitosis classification. 71
- 4-5** Workflow of the cascaded ensemble, which comprises two stages. First, we perform classification with CNN-learned and handcrafted features individually, and if the two classification scores are consistent, a binary decision (mitosis/non-mitosis) will be made directly. Secondly, for those instances whose individual classification scores are highly inconsistent, we classify them again by combining their CNN and handcrafted features. 73
- 4-6** H&E-stained HPF examples from the ICPR dataset. The HPFs are acquired by a Aperio XT scanner with a resolution of  $0.2456\mu m$  per pixel. Each HPF has a size of  $2084 \times 2084$  pixels, representing a  $512 \times 512\mu m^2$  area annotated by pathologists. . . . . 74
- 4-7** Precision-Recall (PR) curve of the proposed HC+CNN method. The performance of the other methods is also plotted for comparison. The curve is generated by varying the classification threshold between 0.45 and 0.7 (The threshold for each point is marked along the curve.). The fact that the performance of the other methods (except IDSIA) lie in the interior of the areas spanned by the curve suggests that the performance of HC+CNN is resilient to the precise choice of the classification threshold. 76

<b>4-8</b>	Mitoses identified by HC+CNN as TP (green rectangles), FN (yellow rectangles), and FP (red rectangles) on the ICPR12 dataset. The TP examples have distinctive intensity, shape and texture while the FN examples are less distinctive in intensity and shape. The FP examples are visually more alike to mitotic figures than the FNs. . . . .	77
<b>4-9</b>	Mitoses identified by CNN and HC+CNN as TP (green circles), FN (yellow circles), and FP (red circles) on a HPF of ICPR12 dataset. On the left side, only using CNN leads to 7 TPs, 5 FNs and 3 FPs. On the right side, using HC and CNN leads to 9 TPs, 3 FNs and 1 FP, which clearly outperforms the use of CNN alone. . . . .	78
<b>4-10</b>	Mitoses identified by HC+CNN as TP (green circles) and FN (yellow circles) on two HPFs of the ICPR12 dataset. Mitoses on the left HPF have distinctive intensities and shapes, and confounding nuclei are few. Therefore, most mitoses can be correctly detected on this HPF. Comparatively, intensity of most mitotic nuclei on the right HPF is not distinctive enough for HC+CNN to identify, as a result, leading to a high FN. . . . .	79
<b>4-11</b>	Overview of the process of training and testing of the deep learning classifiers for invasive breast cancer detection on whole-slide images. . .	82
<b>4-12</b>	(A-C) Example whole-slide images from test TCGA data cohort with ground truth annotations from pathologists, (D-F) the corresponding region predictions produced by the ConvNet classifier and (G-I) region predictions for whole-slide images from the test NC data cohort of normal breast tissue without cancer. . . . .	86
<b>4-13</b>	Example results for the ConvNet <sub>HUP</sub> classifier on the CINJ validation data cohort. The probability map predicted by the ConvNet <sub>HUP</sub> classifier (second row, D-F) was compared against ground truth annotations by a pathologist (first row, A-C). The third row shows the evaluation results of the ConvNet <sub>HUP</sub> classifier in terms of TP (green), FN (red), FP (yellow), and TN (blue) regions. . . . .	88

- 4-14** Whole-slide image from CINJ validation data cohort diagnosed with a rare type of IDC: mucinous carcinoma of the breast. (A) The comparison between the ground truth annotations and the predictions from the ConvNet<sub>HUP</sub> classifier reveal both FN (red) and FP (yellow) errors. (B-C) Most of the FN regions, i.e. tissues wrongly labeled as non-invasive tumor, correspond to mucinous carcinoma, whilst (D) most of FP regions, i.e. tissues wrongly predicted as invasive tumor, are actually invasive mucinous carcinoma that was not included in the annotations by the pathologist. . . . . 89
- 4-15** The most challenging whole-slide image in the CINJ validation cohort achieved the poorest performance via the ConvNet<sub>HUP</sub> classifier with (A) many FP regions and a Dice coefficient of 0.0745. (B) Some of the FN errors are due to the confounding morphologic attributes of the tumor, arising due to a mixing of IDC with fat cells and irregular, infiltrating looking cribriform glands with DCIS. The FP regions appear to be primarily be due to (C) sclerosing adenosis, and (D) DCIS surrounded by IDC. . . . . 90
- 4-16** Agreement plot of the Dice coefficient for the ConvNet<sub>HUP</sub> (X-axis) and ConvNet<sub>UHCMC/CWRU</sub> (Y-axis) classifiers for each slide (blue circles) in the TCGA cohort. The slides with higher disagreement are identified with red circles (see Figure 4-17). . . . . 92
- 4-17** (A-C) Slides from from the TCGA cohort which revealed disagreement between the predictions of the ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> classifiers. The predictions of the (D-F) ConvNet<sub>HUP</sub> and (G-I) ConvNet<sub>UHCMC/CWRU</sub> classifiers were compared against the ground truth annotations in terms of TP (green), FN (red), FP (yellow) and TN (blue) regions. . . . . 93
- 4-18** (A-C) Example whole-slide images from the TCGA data cohort with corresponding ground truth annotations. The probability maps generated by the ConvNet<sub>UHCMC/CWRU</sub> and ConvNet<sub>HUP</sub> classifiers are shown in panels (D-F) and (G-I) respectively. . . . . 94
- 4-19** The probability maps obtained via the ConvNet<sub>UHCMC/CWRU</sub> and ConvNet<sub>HUP</sub> classifiers on whole-slide images of normal breast sections from the UHCMC/CWRU and NC data cohorts are shown in panels (A-C) and (D-F) respectively. 95
- 5-1** Overview of the high-throughput adaptive sampling for whole-slide histopathology images method (HASHI) based on convolutional neural networks (CNN) for automated detection of invasive BCa in WSIs. . . . . 104

<b>5-2</b>	Illustration of the Convolutional Neural Network architecture used to distinguish between invasive BCa and non-invasive BCa. Amongst the various architectures considered, this architecture was found to perform the best. . . . .	107
<b>5-3</b>	Example of 1000 2D-points generated by (a) random (uniform distribution) and quasi-random sampling strategies, (b) Halton and (c) Sobol sequences. (a) We can notice that random generation results in clumping and there are more regions without samples. On the other hand, (b) Halton and (c) Sobol sequences have a more uniform distribution over the whole space with very few clumps, i.e. low discrepancy. . . . .	109
<b>5-4</b>	Comparison between common sampling methods ( <i>regular</i> and <i>dense</i> ) with our novel adaptive sampling method using gradient-based quasi-Monte Carlo sampling ( <i>grad-qmc-halton</i> ). The new unseen WSI (A) with its corresponding ground truth annotation from an expert pathologist (B). The probability maps using regular sampling with step size equal to the patch size (C) and regular dense sampling with step size equal to 1 pixel (D). The new adaptive sampling method shows the iterative process to extract patch samples (E-H) and obtain the probability maps (I-L) for the 1st iteration (E, I), the 2nd iteration (F, J), the 8th iteration (G, K) and the 20th iteration (H, L). . . . .	116
<b>5-5</b>	Quantitative evaluation in a subset of 12 cases from CINJ data cohort among different sampling strategies for invasive BCa detection on WSIs in terms of the Dice coefficient (y-axis) trained with HUP data cohort versus number of samples (x-axis) required for this prediction using a logarithmic scale. . . . .	117
<b>5-6</b>	Performance comparison between $HASHI_{grad-qmc-halton}^{CS256-FC256}$ and $HASHI_{grad-qmc-halton}^{CF-SVM-Linear}$ in terms of Dice coefficient in the independent TCGA test data cohort by varying the classification threshold of the invasive BCa probability map. . . . .	118
<b>5-7</b>	Results of the invasive BCa probability maps (second, fourth and sixth rows) predicted by our CNN-based method $HASHI_{grad-qmc-halton}^{CS256-FC256}$ on representative WSIs from the TCGA test dataset compared to the ground truth annotations from expert pathologists (first, third and fifth rows). . . . .	122



# List of Tables

<b>2-1</b>	Data set distribution per histopathological annotation for training and test. . . . .	30
<b>2-2</b>	Average in automatic annotation performance in both experiments with standard performance measures, Accuracy (Acc), Precision (Pr), Recall (Rc) and F-measure (F) . . . . .	34
<b>2-3</b>	A sample of the visual dictionaries obtained using different local features. Second column shows the 20 visual words with highest probability by class (A:anaplastic, NA:non-anaplastic) for each kind of local feature in a dictionary size of 160. . . . .	43
<b>2-4</b>	Classification performance in terms of accuracy, specificity and sensitivity for dictionary size of 320. . . . .	43
<b>3-1</b>	Classification performance for each learned features from different image datasets using the new DL based method, canonical features, along with different configurations of BOF. The best results are in bold typeface. .	56
<b>3-2</b>	Outputs produced by the system for different cancer and non-cancer input images. The table rows show from top to bottom: the real image class, the input image, the class predicted by the model and the probability associated to the prediction, and the digital stained image (red stain indicates cancer regions, blue stain indicates normal regions). . . .	57
<b>3-3</b>	Medulloblastoma classification performance comparison between unsupervised and supervised feature learning approaches. . . . .	64
<b>4-1</b>	Brief description of handcrafted features used for mitosis detection. . .	72
<b>4-2</b>	Evaluation results for mitosis detection using HC+CNN and comparative methods on the ICPR12 dataset. . . . .	75

<b>4-3</b>	Performances of the blue-ratio segmentation module and the detection module. The blue-ratio segmentation finds 2484 mitosis candidates, among which 93 are true mitoses while the other 2391 are non-mitoses. 7 true mitoses are lost in this step. The detection module identifies 65 true mitoses and 12 false mitoses from these 2484 candidates. 28 mitoses are mis-classified as non-mitotic figures in this module. . . . .	75
<b>4-4</b>	The influence of the number of RF trees. . . . .	77
<b>4-5</b>	Performance measures for the ConvNet classifier on the TCGA (pathological) and NC (normal) data cohorts. The measures included Dice, PPV, NPV, TPR, TNR, FPR and FNR. Note that for the normal cases considered, not all the performance measures are shown because the NC data cohort did not have cancer annotations. . . . .	85
<b>4-6</b>	Performance of the ConvNet <sub>HUP</sub> and ConvNet <sub>UHCMC/CWRU</sub> classifiers on the CINJ data cohort in terms of means and standard deviation of Dice coefficient, PPV and NPV. The results in Table <b>4-6</b> are organized in terms of all cases in the CINJ cohort (N=40), a subset of the CINJ cohort with invasive breast cancer alone (N=19), and a mixture of invasive and other in situ subtypes of breast cancer (N=21). . . . .	87
<b>4-7</b>	Comparison and correlation of the ConvNet <sub>UHCMC/CWRU</sub> and ConvNet <sub>HUP</sub> classifiers in terms of Dice, PPV, NPV, TPR, TNR, FPR and FNR. Note that for the normal cases considered, not all the performance measures are shown because the NC data cohort did not have cancer annotations. . . . .	91
<b>5-1</b>	Breast cancer data cohorts used for training, validation and testing in the experimental evaluation. . . . .	110
<b>5-2</b>	Set of hand-crafted features used for comparison against the CNN based feature learning approach. . . . .	120
<b>5-3</b>	Classification performance comparison in the CINJ validation dataset among different architectures of CNN models and state-of-the-art hand-crafted features trained with HUP dataset and evaluated in terms of AUC. . . . .	121
<b>5-4</b>	Invasive BCa detection performance of our HASHI method on the testing data cohort of TCGA in terms of Dice, PPV, NPV, TPR, TNR, FPR and FNR. . . . .	121

# 1 Introduction

The primary goal of this thesis is to address the problem of analyzing histopathology images, using computational tools, to automatically find patterns related with pathology signatures associated to healthy and abnormal tissues, which are a fundamental support for cancer diagnosis. Computational pathology is a relatively recent research area devoted to providing accurate and efficient computational methods to support quantitative detection, diagnosis, and prognosis in pathology. This work presents several computational learning methods and frameworks for automatic histopathology image representation learning from histopathology image databases for different digital pathology tasks including tumor and tissue detection, location and quantification in several cancer types.

This introduction chapter includes the motivation for this work followed by the statement and definition of the research problem and a description of the computational challenges faced in this thesis. The chapter also includes a review of the current state of the art on representation learning for digital pathology image analysis. Finally, the contributions and academic products developed during this thesis are presented, followed by the description of thesis document organization.

## 1.1 Motivation

Cancer research is a major public health priority in the world due to its high incidence, diversity and mortality. Cancer is the common name used to identify a collection of related diseases which involve abnormal cell growth with the potential to invade or spread to other parts of the body. Hence, there are more than 100 different types of cancer depending on cell type origin and grade of aggressiveness. Particularly, the types of cancers addressed in this work include those of the skin, brain and breast:

- Skin cancer has gradually become more common worldwide and in white populations is the most common malignancy. In Colombia, skin cancer comprises 30% of the total cancer diagnosed [52]. About 80% of skin carcinomas that are not melanoma corresponds to basal cell carcinoma whereas the remaining 20%

of cases are squamous cell carcinoma [2, 143]. Despite not being the most deadly skin cancer, basal cell carcinoma is the most common and its incidence is growing worldwide [65].

- Brain cancer and other nervous system cancer have been diagnosed on approximately 0.6% of men and women at some point during their lifetime. This represents 2.4% of cancer deaths in United States. In Colombia, the deaths caused by malignant brain cancer is around 1.5 per each 100,000 inhabitants. Medulloblastoma brain cancer is one of the most common types of malignant brain tumors [141]. In adults, the disease is rare whereas in children the incidence amounts to a 25% of all pediatric brain tumors. Different histologic types of medulloblastoma have different prognoses. For instance, anaplastic medulloblastomas have worse prognosis than non-anaplastic variants. The therapeutic management changes radically depending on the subtype of medulloblastoma; hence, histopathological diagnosis is useful in determining the potential outcome of the disease.
- Breast cancer is the most common type of cancer in women and the second cause of death in developed countries [125, 148], with the American Cancer Society estimating that 1 in 8 women develop breast cancer during their life in United States [1]. In Colombia, this is the main cause of death by cancer in women. Invasive breast cancers are those that spread from the original site (either the milk ducts or the lobules) into the surrounding breast tissue. These kinds of breast cancer comprises about 70% of cases [48, 53] and they have poorer prognosis than in-situ types [53]. Isolation of invasive breast cancer allows further analysis of tumor differentiation via the Bloom-Richardson and Nottingham grading schemes, which estimate cancer aggressiveness by evaluating histologic characteristics including: tubule formation, nuclear pleomorphism and mitotic count [71].

According to World Health Organization estimates for 2011, cancer now causes more deaths than all coronary heart diseases or all strokes. The continuing global demographic and epidemiologic transitions indicate an ever-increasing cancer burden over the next decades, particularly in low and middle income countries, with over 20 million new cancer cases expected annually as early as 2025. The last report of cancer incidence of prevalence and mortality from GLOBOCAN project, reveals that there were 14.1 million new cases and 8.2 million deaths in 2012, where 60% of world's total new annual cases occur in Africa, Asia and Central and South America and 30% of those cancers could be prevented [64].

Despite the great advances in cancer research, diagnosis and treatment in recent decades [13, 146], the high incidence and lack of specialists have proven that one of

the major challenges is to achieve early diagnosis [79]. The possibilities of timely treatment and the survival chances of patients can be improved with an early diagnosis, especially in developing countries such as Colombia [63]. The development of new technologies to support cancer detection and diagnosis could help to reduce the workload for expert pathologists by triaging relatively simple cases and allowing them to spend additional time focusing on the more challenging cases, potentially speeding up the screening process in the early stages.

Histopathology is the study of cell anatomy and tissues to determine the presence of elements associated to a particular pathology or disease in a microscopic level with diagnosis purposes. Histopathology images are very important for diagnosis purposes. These images are a fundamental resource to determine the state of a particular biological structure, to support diagnosis of diseases like cancer, or to analyze the state of organs and tissues [6]. Pathologists are trained in histopathology to interpret the appearance of tissues according to their structure, functionality and cellular organization at different organs, those features can be highlighted using several staining processes. This kind of images are used for both cancer research and for making clinical decisions, and are mainly used as diagnosis ground truth for other screening studies such as X-ray and magnetic resonance imaging [68, 78].

Thanks to recent advances in scanning technology for pathology slides digitization and the growth of global initiatives to build databases for cancer research, such as The Cancer Genome Atlas <sup>1</sup>, *digital pathology* has emerged as a new approach to improve and support routine clinical and research pathology workflows, based on the analysis and diagnosis of digitalized whole-slide histopathology images by pathologists [113]. Typically, a digital pathology image analysis workflow consists of the following stages (see Figure 1-1): i) image preprocessing for color normalization, ii) automatic detection and segmentation of biological structures (e.g. nuclei, cells, tissues), iii) feature extraction algorithms to capture color-based, morphological, topographical and multi-scale visual patterns, and iv) classification or quantification of biomedical or diseases signatures as diagnosis support [68, 73].

Digital pathology has boosted the development of several computational methods, proposed by the image processing, machine learning and computer vision communities. Some applications include nuclei and cell detection [44, 120], morphological and architectural cell arrangement characterization [9, 58], tissue type classification [51, 69, 132], tumor detection [121], aggressiveness estimation and grading [10, 44] and patient outcome estimation [14, 24, 157].

In fact, thanks to the usage of computerized methods for histopathology image analysis, such as was described above, and the analysis of public available databases, currently

---

<sup>1</sup>The Cancer Genome Atlas: <http://cancergenome.nih.gov/>

known as *in-silico* analysis, some remarkable works have been published in the literature with interesting contributions and findings. For instance, Beck et al. [14] found that some stromal features are significantly associated with survival in breast cancer, and this association was stronger than the association of survival with epithelial characteristics of the C-path (Computational pathologist) system. On the other hand, there are other works using the conventional histopathology image analysis framework to identify image-based features associated with molecular or genetic cancer subtypes for outcome and survival prediction [24, 157].

Despite the fact that most of these approaches had shown its capability and potential for supporting diagnosis in pathology for different cancer types and pathology tasks, still there are some limitations and drawbacks for their application in clinical practice. Among the major limitations is the development of accurate, robust and time-efficient methods applicable on independent data sources in related tasks over different cancer types, which could be able to guarantee quantifiable objectivity in terms of confidence and reproducibility.

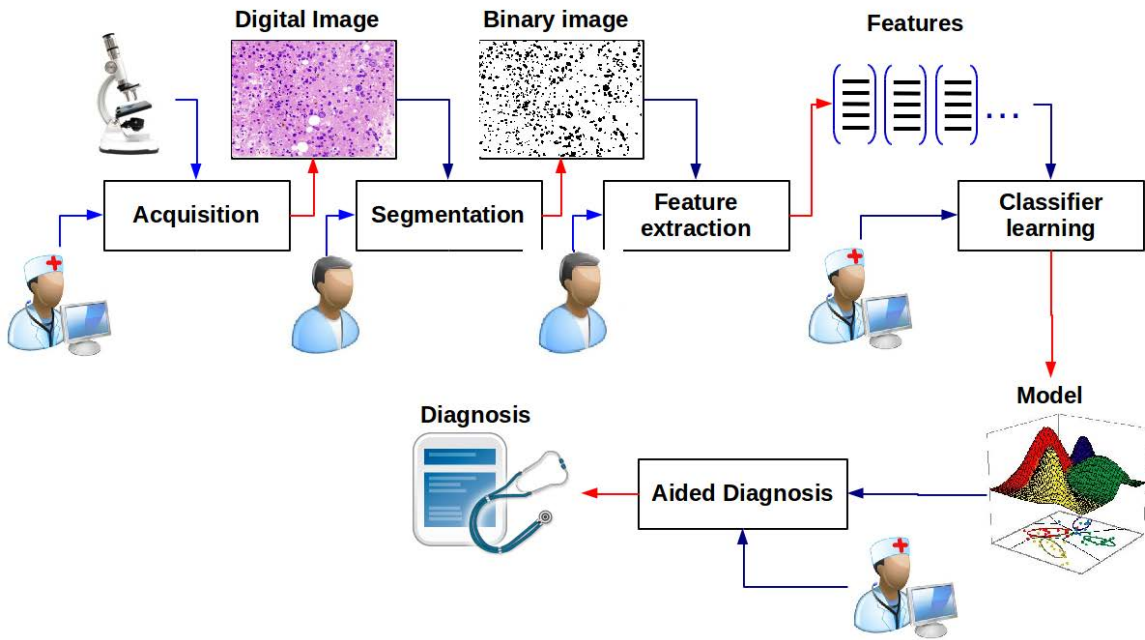


Figure 1-1: Common digital pathology workflow.

On the other hand, machine learning techniques have focused on exploiting large databases (a.k.a. *Big Data*) for designing data-driven automatic learning methods. Several

new approaches have emerged in opposition to traditional schemes based on combinations of manually engineered (or hand-crafted) features and machine learning classifiers, which nowadays are part of an area known as *representation learning*. Representation learning is related to learn transformations of the data that make easier the extraction of useful information when building classifiers or other predictors [16]. This definition includes most of the data-driven approaches used in the last years, which have been leading the state of the art in most pattern recognition tasks until now, including: bag of features [43, 109, 128], dictionary learning [115, 153], sparse representations [110, 115, 130, 161], matrix factorization methods [54, 55, 103, 104], and deep learning [16, 145].

Thanks to the growth of diverse whole-slide histopathology image databases and the convergence of both trends, digital pathology and representation learning, several successful works have been proposed for different pathology tasks, such as nuclei and mitosis detection [28, 87, 112, 156, 159], tissue classification [24, 31, 51, 69, 132, 139], tumor detection and grading [7, 121, 140], with very promising results.

Exploiting the increasing number of databases in digital pathology through data-driven machine learning approaches, such as representation learning methods, there could be a breaking point in how cancer diagnosis and research are performed in current pathology workflows. It is important that these methods can be time-efficient, robust, reproducible, accurate and with quantifiable performance to support the decision making of pathologists, in terms of early cancer detection and diagnosis, tumor grading, prognosis and theragnosis, all of these tasks fundamental to evolving towards personalized medicine.

## 1.2 Research Problem

In contrast with the annotations of visual patterns and concepts in natural images which are associated with subjects, objects and actions [46], histopathology image annotations are related to pathological lesions, morphological and architectural features, which encompass a complex mixture of visual patterns that allow to decide about the illness presence. In general, histopathology images with the same annotations present a high visual variability, which can be generated by several factors, starting with the inevitable uncertainty coming from the fact that a very complicated 3D biological structure is randomly projected onto a 2D image. For achieving so, tissues are subjected to a series of histological procedures: they are initially fixated with a basic aldehyde, then dehydrated, embedded and finally cut [6] (Figure 1-2). This chain of events is very susceptible to different kinds of cumulative errors that result in histopathological images with a complex mix of patterns and sub-patterns which are interpreted by an



Figure 1-2: Histotechnique procedure and digitalization process. From top left to bottom right: 1) Small cuts are made from the tissue 2-3) The smaller cuts are put into cassettes and (after processing) embedded in paraffin 4) Thin sections are made from the paraffin blocks with a microtome 5) The sections are put onto glass slides for staining 6) H&E stained slide which is 7) digitized with a pathology slide scanner to obtain 8) a whole-slide image. (Adapted from: <http://amida13.isi.uu.nl/?q=node/2>)

expert, even in cluttered biological circumstances. In addition, image capturing parameters such as environment illumination, exposure time, microscope magnification, etc., are a source of image variability. Finally, the ground truth annotations from expert pathologists have a subjectivity component that produces a certain degree of inter- and intra- variability for diagnosis or pattern location [61].

Therefore, the relevant visual pathological patterns highly change their visual appearance according to their spatial location, severity and co-occurrence with others biological structures (Figure 1-3). Hence, considering the above reasons, histopathology image databases are highly variable and susceptible to yield on *batch effects*, i.e., non-biological experimental variations such as age of sample, method of slide preparation, staining, specifications of the imaging device, and type of post-processing software [90, 137].

A characteristic feature of digital pathology is the large volume of data produced. There are two main different causes for this phenomenon. First, whole-slide images (WSI) in histopathology are large, typically each WSI could have a spatial resolution of  $80,000 \times 80,000$  pixels and have a size around 20 GB [83, 92]. Second, the emerging of world-wide initiatives in cancer research had been collecting large data bases including



different source of biomedical data such as clinical and pathology reports, digitized whole-slide histopathology images and several types of genomic data. Examples of this are the National Cancer Institute (NCI) Cooperative Prostate Cancer Tissue Resource, the National Institute of Health (NIH) Cancer Genome Atlas, the Human Protein Atlas [118,119,154], and the Cancer Genome Atlas (TCGA) [126]. The TCGA started in 2005 and it had collected around 11,079 cases from 34 different types of cancer. In fact, TCGA project is a good example of big data in cancer research involving different types of data (clinical, pathological and genomic). The TCGA now has an approximate size of 1,095 Terabytes (1 Petabyte) comprising more than 79,000 files and growing at the rate of 10 Terabytes per year [74].

Considering the above statements, translational medicine in histopathology is still an open problem. Despite that there are several works designing computerized methods for image analysis in digital pathology tasks [6,73,106,155] and significant advances in high-speed and high-resolution scanning devices, there are not enough works tackling the robustness, reproducibility and feasibility of those computerized solutions in whole-slide images for real clinical and research scenarios [81,83,91,92,147]. The two main reasons are: first, the few public databases from different institutions or pathology labs in several cancer types to evaluate the reproducibility and robustness of computerized methods in different patient cohorts and digital pathology tasks, and second, the large size and volume of whole-slide images with heterogeneous tissues, which require not only accurate but also high-throughput, time-efficient computerized methods [68,92,113].

### 1.2.1 Research Question

Despite the fact that there are several computerized methods applied to different pathology tasks, few of them focus on extracting the best underlying data structure that represent the meaningful histopathology patterns through feasible, efficient, robust and reproducible solutions to support digital pathology workflows in real clinical and research scenarios. Nowadays, this goal becomes more relevant considering the advancements of digital pathology, the growth of public databases with thousands of large size whole-slide images and the rise of data-driven methods based on machine learning methods. Thus, these considerations have led us to the main research question in this work: *how to efficiently learn an appropriate data representation directly from histopathology image databases that can be effective in different digital pathology tasks?*

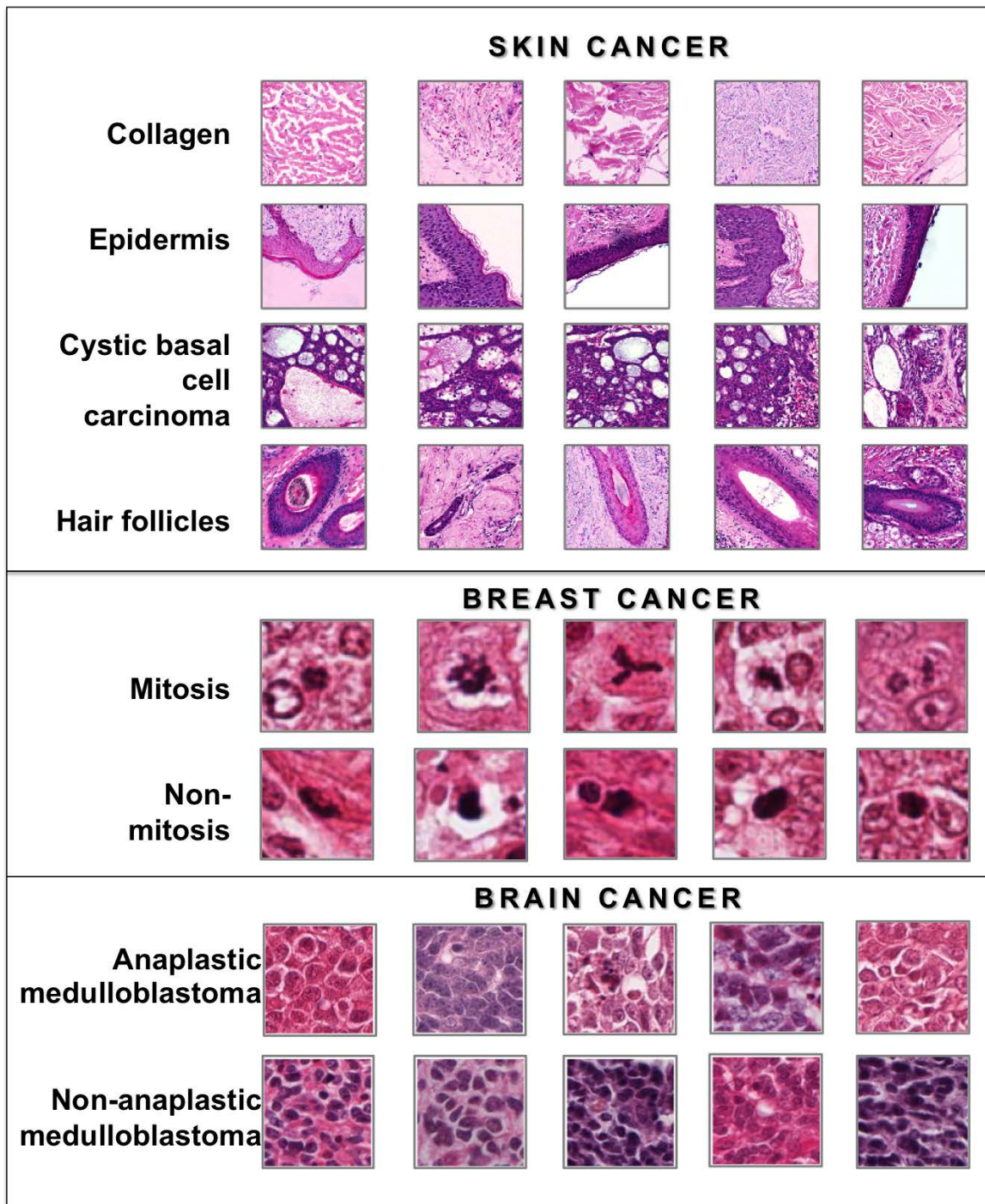


Figure 1-3: Visual variability of several normal and pathological patterns in histopathology images in different cancer types (skin, breast and brain).

## 1.3 Challenges

Satisfactorily answering this question poses different technical challenges, which were addressed in this thesis, and are described next:

- To tackle the high visual variability of histopathology images by adapting or proposing successful and accurate state-of-the-art computational learning methods that learn directly from data.
- To take advantage of growing histopathology image databases available to learn appropriate and useful representations of the data that could be reproducible for several digital pathology tasks.
- To devise robust learning algorithms that can be applicable to different cancer types, variable and noisy data sources and heterogeneous digital pathology goals.
- To achieve accurate and efficient computational learning models for prediction, which can obtain comparable results with ground truth annotations from expert pathologists in feasible computing time.
- To deal with the large size resolution and volume storage of whole-slide images for high-throughput image analysis in terms of efficient or feasible computing time.

## 1.4 Contributions and Academic Products

This work presents several contributions to digital pathology image analysis through computational machine learning approaches supported by several histopathology databases and high-throughput computing. The main contributions of this work are data-driven representation learning frameworks for automatic histopathology image analysis to support diagnosis in different cancer types. These methods are focused to provide accurate, robust, reproducible and time-efficient tools that can be applicable in clinical practice and cancer research. The different publications done as a part of this thesis are listed as follows organized according to the main biomedical application:

### 1.4.1 Semantic Part-based image Representations

#### Classification of Basal Cell Carcinoma in Skin Cancer

- **Angel Cruz-Roa**, Gloria Díaz, Eduardo Romero, Fabio González. *Automatic Annotation of Histopathological Images Using a Latent Topic Model Based On*

*Non-negative Matrix Factorization*. Journal of Pathology Informatics. 2011; 2:4. ISSN: 2153-3539. doi:10.4103/ 2153-3539.92031

- **Angel Cruz-Roa**, Juan C. Caicedo, Fabio A. González. *Visual Pattern Mining in Histology Image Collections Using Bag of Features*. Journal Artificial Intelligence in Medicine. Vol 52 (2011) pp. 91-106. ISSN: 0933-3657, doi:10.1016/j.artmed.2011.04.010
- A. Pastor López-Monroy, Manuel Montes-y-Gómez, Hugo Jair Escalante, **Angel Cruz-Roa**, Fabio A. González. *Bag-of-Visual-Ngrams for Histopathology Image Classification*. Proc. SPIE 8922, IX International Seminar on Medical Information Processing and Analysis - SIPAIM 2013, 89220P. Mexico D.F. (Mexico). 11-14 Noviembre (2013). doi:10.1117/12.2034113
- **Angel Cruz-Roa**, Gloria Díaz, Eduardo Romero, Fabio González. *Automatic Annotation of Histopathological Images Using a Latent Topic Model Based On Non-negative Matrix Factorization*. MICCAI 2011, the 14th International Conference on Medical Image Computing and Computer Assisted Intervention - Histopathology Image Analysis Workshop. September 18-22 (2011)
- **Angel Cruz-Roa**, Gloria Díaz, Fabio González. *A Framework for Semantic Analysis of Histopathological Images Using Nonnegative Matrix Factorization*. Computing Congress (CCC), 2011 6th Colombian. May 4-6 2011. Universidad Nacional de Colombia, Sede Manizales. Manizales (Colombia). ISBN: 978-1-4577-0285-3, doi:10.1109/COLOMCC.2011.5936285

### Classification of Medulloblastoma Tumor in Brain Cancer

- **Angel Cruz-Roa**, Fabio González, Joseph Galaro, Alexander R. Judkins, David Ellison, Jennifer Baccon, Anant Madabhushi, Eduardo Romero. *A Visual Latent Semantic Approach for Automatic Analysis and Interpretation of Anaplastic Medulloblastoma Virtual Slides*. MICCAI 2012, the 15th International Conference on Medical Image Computing and Computer Assisted Intervention. Nice (France), October 1-5 2012. Lectures Notes in Computer Sciences. Vol. 7510. pp 157-164. ISSN 0302-9743. doi:10.1007/978-3-642-33415-3\_20

## 1.4.2 Feature and Representation Learning

### Mitosis Detection in Breast Cancer

- Haibo Wang, **Angel Cruz-Roa**, Ajay Basavanhally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio González, and Anant Madabhushi. *Mitosis Detection in Breast Cancer Pathology Images by Combining Handcrafted and Convolutional Neural Network Features*. Journal of Medical Imaging. 1(3):034003 (2014). ISSN: 2329-4302. doi:10.1117/1.JMI.1.3.034003
- Mitko Veta, Paul J. van Diest, Stefan M. Willems, Haibo Wang, Anant Madabhushi, **Angel Cruz-Roa**, Fabio González, Anders B. L. Larsen, Jacob S. Vestergaard, Anders B. Dahl, Dan C. Cireşan, Jürgen Schmidhuber, Alessandro Giusti, Luca M. Gambardella, F. Boray Tek, Thomas Walter, Ching-Wei Wang, Satoshi Kondo, Bogdan J. Matuszewski, Frederic Precioso, Violet Snell, Josef Kittler, Teofilo E. de Campos, Adnan M. Khan, Nasir M. Rajpoot, Evdokia Arkoumani, Miangela M. Lacle, Max A. Viergever, Josien P.W. Pluim *Assessment of algorithms for mitosis detection in breast cancer histopathology images*. Journal of Medical Image Analysis. Volume 20, Issue 1, Pages 237-248. 2015. ISSN:1361-8415. doi: 10.1016/j.media.2014.11.010
- Haibo Wang, **Angel Cruz-Roa**, Ajay Basavanhally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio González, and Anant Madabhushi. *Cascaded Ensemble of Convolutional Neural Networks and Handcrafted Features for Mitosis Detection*. Digital Pathology Conference. SPIE Medical Imaging 2014, 15 - 20 February 2014. Town & Country Resort and Convention Center, San Diego, California, USA. doi:10.1117/12.2043902

### Invasive Tumor Detection in Breast Cancer

- **Angel Cruz-Roa**, Hannah Gilmore, Ajay Basavanhally, Michael Feldman, Shridhar Ganesan, Natalie Shih, John Tomaszewski, Fabio González, Anant Madabhushi. *Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning based tool for quantifying tumor location and extent*. Submitted to Journal of the American Medical Informatics Association. 2015.
- **Angel Cruz-Roa**, Ajay Basavanhally, Fabio Gonzalez, Michael Feldman, Shridhar Ganesan, Natalie Shih, John Tomaszewski, Hannah Gilmore, and Anant Madabhushi. *A Feature Learning Framework for Reproducible Invasive Tumor Detection of Breast Cancer in Whole-Slide Images*. USCAP 104th Annual Meeting, In Laboratory Investigation (Vol. 95, pp. 40A-40A). March 21-27, 2015 in

Boston, MA. 75 Varick ST, 9TH FL, New York. USA: Nature Publishing Group. (Abstract)

- **Angel Cruz-Roa**, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski and Anant Madabhushi. *Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks*. Digital Pathology Conference. SPIE Medical Imaging 2014, 15 - 20 February 2014. Town & Country Resort and Convention Center, San Diego, California, USA. doi:10.1117/12.2043872

### Classification of Basal Cell Carcinoma in Skin Cancer

- John Arévalo, **Angel Cruz-Roa**, Viviana Arias, Eduardo Romero, Fabio A. González. *An unsupervised feature learning framework for basal cell carcinoma image analysis*. Journal of Artificial Intelligence in Medicine, Available online 23 April 2015, ISSN 0933-3657. doi:10.1016/j.artmed.2015.04.004 (In press)
- John Arévalo, **Angel Cruz-Roa** and Fabio A. González. *Hybrid image representation learning model with invariant features for basal cell carcinoma detection*. Proc. SPIE 8922, IX International Seminar on Medical Information Processing and Analysis - SIPAIM 2013, 89220M. Mexico D.F. (Mexico). 11-14 Noviembre (2013). doi:10.1117/12.2035530
- **Angel Cruz-Roa**, John Arévalo, Anant Madabhushi, Fabio González. *A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection*. MICCAI 2013, the 16th International Conference on Medical Image Computing and Computer Assisted Intervention. Nagoya (Japan), September 22-26 2013. Lectures Notes in Computer Sciences. Vol. 8150. pp 403-410. (Student Travel Award)

### Classification of Medulloblastoma Tumor in Brain Cancer

- **Angel Cruz-Roa**, John Arévalo, Ajay Basavanhally, Anant Madabhushi, Fabio González. *A comparative evaluation of supervised and unsupervised representation learning approaches for anaplastic medulloblastoma differentiation*. Proc. SPIE 9287, Tenth International Symposium on Medical Information Processing and Analysis (SIPAIM 2014), Cartagena, Colombia. October 14-16, 2014. 92870G (January 28, 2015); doi:10.1117/12.2073849

### 1.4.3 Accurate and high-throughput whole-slide histopathology image analysis

#### Invasive Tumor Detection in Breast Cancer

- **Angel Cruz-Roa**, Ajay Basavanahally, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, Hannah Gilmore, Anant Madabhushi and Fabio González. *Scaling up convolutional networks for invasive breast cancer detection in whole-slide histopathology images through high-throughput adaptive sampling*. Submitted to Journal IEEE Transactions on Medical Imaging. 2015.

### 1.4.4 Other contributions

Additional papers were published as a result of collaborations performed in different and related research areas.

#### Histopathology image analysis

- Sebastian Otálora, **Angel Cruz-Roa**, John Arévalo, Manfredo Atzori, Anant Madabhushi, Alexander R. Judkins, Fabio González, Henning Müller, Adrien Depeursinge. *Anaplastic medulloblastoma tumor differentiation by combining unsupervised feature learning and Riesz wavelets for histopathology image representation*. MICCAI 2015, the 18th International Conference on Medical Image Computing and Computer Assisted Intervention. Munich (Germany), October 5-9 2015. (Accepted)
- David Romo, Ricardo Moncayo, **Angel Cruz-Roa**, Eduardo Romero. *Identifying Histological Concepts on Basal Cell Carcinoma Images using Nuclei based Sampling and Multi-Scale Descriptors*. Biomedical Imaging: From Nano to Macro, ISBI'2015: Proceedings of the 9th IEEE International Symposium on. April 16-19 2015. New York, NY, USA.
- John Arévalo, **Angel Cruz-Roa**, Fabio A. González. *Histopathology image representation for automatic analysis: A state-of-the-art review*. Revista Med. Vol. 22 (2), pp. 79-91. 2014. ISSN: 0121-5256
- **Angel Cruz-Roa**, Jun Xu, Anant Madabhushi. *A note on the stability and discriminability of graph based features for classification problems in digital pathology*. Proc. SPIE 9287, Tenth International Symposium on Medical Information Processing and Analysis (SIPAIM 2014), Cartagena, Colombia. October 14, 2014. 928703 (January 28, 2015); doi:10.1117/12.2085141.

- **Angel Cruz Roa**, Eduardo Romero Castro, Fabio González Osorio. *An adaptive image representation learned from data for cervix cancer tumor detection*. Digital Pathology Conference. SPIE Medical Imaging 2013, Volume 8676. 9 - 14 February 2013. Disney's Coronado Springs Resort, Lake Buena Vista (Orlando Area), Florida USA (*Poster Award Cum Laude*).
- **Angel Cruz-Roa**, John Arévalo, Fabio González. *Prediction of Morphometric Measures from Bag-of-Features Image Representation of Cervix Cancer Cells*. 8th International Seminar on Medical Information Processing and Analysis - SIPAIM 2012. San Cristobal (Venezuela). 12-15 Noviembre (2012)

### Radiology image analysis

- Andrea Rueda, John Arévalo, **Angel Cruz-Roa**, Eduardo Romero, Fabio A. González. *Bag of Features for Automatic Classification of Alzheimer's Disease in Magnetic Resonance Images*. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. The 17th Iberoamerican Congress on Pattern Recognition (CIARP) Lecture Notes in Computer Science Volume 7441, 2012, pp 559-566. September 3-6, Buenos Aires (Argentina). 2012. ISSN 0302-9743.
- **Angel Cruz-Roa**, Angélica Atehortúa, Fabio González. *An evaluation of different image features for modality classification in ImageClefmed 2012 dataset*. 8th International Seminar on Medical Information Processing and Analysis - SIPAIM 2012. San Cristobal (Venezuela). 12-15 Noviembre (2012).
- **Angel Cruz-Roa** and Fabio González. *Multi-label Annotation of Radiological Images Based On Visual Latent Semantic Analysis*. 7th International Seminar on Medical Information Processing and Analysis -SIPAIM 2011. 5-7 December 2011. Universidad Industrial de Santander. Bucaramanga (Colombia).

### Lung sounds detection

- Germán Sosa, **Angel Cruz-Roa**, Fabio González. *Automatic detection of Wheezes by evaluation of multiple acoustic feature extraction methods and C-weighted SVM*. Proc. SPIE 9287, Tenth International Symposium on Medical Information Processing and Analysis (SIPAIM 2014), Cartagena, Colombia. October 14-16, 2014. 928709 (January 28, 2015); doi:10.1117/12.2073614
- Germán Sosa, **Angel Cruz-Roa**, John Arévalo, Fabio González. *An unsupervised feature learning approach based on denoising autoencoders for wheeze lung*



*sounds detection*. 2015. (In preparation)

### High-performance computing for image analysis

- Raúl Ramos-Pollán, Fabio González, Juan C. Caicedo, **Angel Cruz-Roa**, Jorge E. Camargo, Jorge A. Vanegas, Santiago A. Pérez, Jose David Bermeo, Juan Sebastián Otálora, Paola K. Rozo, John E. Arévalo. *BIGS: A Framework for Large-Scale Image Processing and Analysis Over Distributed and Heterogeneous Computing Resources*. In IEEE International Conference on eScience. October 8-12, Chicago (USA). 2012.
- Raúl Ramos-Pollán, **Angel Cruz-Roa**, Fabio González. *A Framework for High Performance Image Analysis Pipelines over Cloud Resources*. The Latin American Conference on High Performance Computing (CLCAR), Panamá 2012.
- Raúl Ramos-Pollán, **Angel Cruz-Roa**, Fabio González. *A Framework for High Performance Image Analysis Pipelines*. 7th Colombian Computing Congress (7CCC) 2012.
- Paola K. Rozo, **Angel Cruz-Roa**, Jorge A. Vanegas, Jorge E. Camargo, Juan C. Caicedo, Fabio A. González. *Sistema de almacenamiento, recuperación semántica y anotación automática de imágenes histológicas en una infraestructura Grid por medio de RENATA*. Revista e-colabora. Vol 2, No 4 (2012). ISSN: 2027 - 7415.

## 1.5 Organization of this thesis

The remaining thesis chapters are organized as follows:

- **Chapter 2: Automatic Annotation of Histopathological Images with a NMF-based Latent Topic Model.** This chapter presents a new method for automatic annotation of morphological and architectural patterns on histopathology images. The proposed method is an annotation framework using a latent topic model based on non-negative matrix factorization which learned the relationship between the visual content of an image represented as bag of features and its high-level concept meaning. In addition to the annotation framework, the method includes interpretation support in its results by providing probability values for each annotation class and a visual map highlighting the corresponding regions. The method was validated in two different scenarios: identification of normal and cancerous tissues in basal-cell carcinoma skin cancer and tumor differentiation of medulloblastoma brain cancer. In both scenarios, the proposed

approach outperformed previous and related works from the state of the art, and also improved the interpretability of the results.

- **Chapter 3: *Unsupervised histopathology image representation learning.*** This chapter presents novel approaches based on unsupervised representation learning using a different kind of auto-encoders in two different digital pathology tasks for histopathology image analysis: basal-cell carcinoma tumor detection in skin cancer and tumor differentiation of medulloblastoma brain cancer. In both tasks, auto-encoders with different characteristics were evaluated including sparse auto-encoders, reconstruct independent component analysis and topographic reconstruct independent component analysis. Each method was trained in a totally unsupervised manner directly from tissue samples of histopathology images and then a supervised training was performed by using a softmax classifier for each particular classification task. In addition, a new method to provide interpretable support was developed, which works as a *digital staining* that highlights the spatial regions associated to cancerous tissues. The best results that improved the state of the art in both tasks were achieved by the topographic reconstruct independent component analysis auto-encoders, which learned features with invariance properties in scale and orientation that captured the complexity of histopathology patterns.
- **Chapter 4: *Supervised histopathology image representation learning.*** This chapter presents novel approaches based on supervised representation learning using convolutional neural networks in two different digital pathology tasks for histopathology image analysis in breast cancer: mitosis and invasive tumor detection. For mitosis detection, a combined approach between handcrafted features from the state of the art and learned features from a 3-layer convolutional neural network was able to obtain good results in the challenging task of mitosis detection in high-resolution histopathology images with efficient computing time. For invasive breast cancer detection, a shallow convolutional neural network was trained from manual annotations of pathologists of invasive tumor regions to predict a probability map of invasive breast cancer regions in large whole-slide histopathology images. In both tasks, the convolutional neural networks were trained and validated with independent data cohorts and compared with ground truth annotations from expert breast cancer pathologists. The approaches show good performance in two challenging and relevant tasks for breast cancer diagnosis. Firstly in mitosis detection, one of the three criteria to determine the tumor grade according to Bloom-Richardson and Nottingham grading schemes. Secondly in invasive breast cancer detection, which is around of 80% of breast

cancer cases that comprises the more aggressive forms (invasive ductal carcinoma and invasive lobular carcinoma), where the accurate and fast detection is determinant for posterior analysis of difficult cases or tumor grade estimation.

- **Chapter 5: *High-throughput whole-slide histopathology image analysis*.** This chapter presents a new high-throughput adaptive sampling method for accurate whole-slide histopathology image analysis. Our method includes a quasi-Monte Carlo sampling and a gradient-based uncertainty mapping for iteratively and adaptively take more samples in higher uncertain regions and less samples in confident areas with only a type of tissue (invasive or not). Each tissue sample is classified by an accurate supervised feature learning classifier based on convolutional neural networks, which provided a probability between zero and one of belonging to an invasive breast cancer tissue. The method was validated in the task of automatic region detection of invasive tumors in larger whole-slide images of estrogen receptor positive patients of breast cancer. The evaluation was done by comparing the predicted regions of our method and the ground truth annotations from pathologists to measure the accuracy in different and independent data cohorts including a subset of the public and well known database of *The Cancer Genome Atlas*. The proposed tissue classification method based on convolutional networks was evaluated and compared against the state-of-the-art handcrafted features for histopathology image analysis obtaining the best results. The high-throughput in computing time of our method was compared with traditional and conventional Monte Carlo sampling methods achieving the best performance in terms of accuracy and time for whole-slide image analysis.
- **Chapter 6: *Conclusions and Perspectives*.** This final chapter presents the main conclusions of this thesis, highlighting the main contributions, the most important findings and their impact in research and practical areas. Finally, future research directions and perspectives are presented and discussed.

## **Part I**

# **Semantic Part-based image Representations**

## 2 Automatic Annotation of Histopathological Images with a NMF-based Latent Topic Model

*The first step in histopathology image analysis is to identify morphological and architectural features from nuclei, cell and tissues that could be associated with pathological or normal patterns. These histopathology features are very important for posterior differential diagnosis to determine the patient condition, either healthy or pathological including cancer. This chapter presents a novel method for representation, annotation and interpretable visual analysis, through semantic digital staining, of histopathology images, and its application in skin (see into Section 2.1) and brain cancer (see into Section 2.2) histopathology images. The following sections correspond to papers presenting the details of the methods and applications.*

### 2.1 Histopathology Image Annotation in Basal-cell Carcinoma

Histopathological images are an important resource for clinical diagnosis and biomedical research. From an image understanding point of view, the automatic annotation of these images is a challenging problem. This paper presents a new method for automatic histopathological image annotation based on three complementary strategies, first, a part-based image representation, called the bag of features, which takes advantage of the natural redundancy of histopathological images for capturing the fundamental patterns of biological structures, second, a latent topic model, based on non-negative matrix factorization, which captures the high-level visual patterns hidden in the image, and, third, a probabilistic annotation model that links visual appearance of morphological and architectural features associated to 10 histopathological image annotations. The method was evaluated using 1,604 annotated images of skin tissues, which included normal and pathological architectural and morphological features, obtaining a recall of 74% and a precision of 50%, which improved a baseline annotation method based

on support vector machines in a 64% and 24%, respectively. *The complete content of this section has been published as a research article in the **Journal of Pathology Informatics**, and previous works had been published in **MICCAI 2011**, the **14th International Conference on Medical Image Computing and Computer Assisted Intervention - Histopathology Image Analysis Workshop** and **2011 6th Colombian Computing Congress (CCC)** (see [30, 39, 40]).*

### 2.1.1 Introduction

Recent advances in microscopical acquisition technology have allowed to collect huge numbers of histopathological images, an important resource for the diagnosis act as well as for pathologist training [113]. The interest in developing the suitable image technology to address the automatic analysis of this kind of images has rapidly grown over the last years [94, 107, 151]. As a consequence, a new research area, called bioimage informatics, has emerged integrating data mining, database visualization, extraction, searching, comparison and management of biomedical visual data [135, 151]. This area combines both image analysis and computational techniques to provide powerful tools that facilitate high-throughput/high-content analysis of biological tissues [135].

Automatic annotation of histopathological images is a very challenging problem. In contrast with natural images, high level annotations are not usually associated to particular objects in the image. In histopathological images, annotations are related to pathological lesions, morphological and architectural features, which encompass a complex mixture of visual patterns that allow to decide about the illness presence. In general, images with the same annotations present a high visual variability, which can be generated by several factors, starting with the inevitable uncertainty coming from the fact that a very complicated 3D biological structure is randomly projected onto a 2D image i.e. the tissue must become a solid structure from which surface cuts of barely  $1 - 5 \mu\text{m}$  are obtained. For achieving so, tissues are subjected to a series of histological procedures: they are initially fixated with a basic aldehyde, then dehydrated, embedded and finally cut. This chain of events is very susceptible to different kinds of cumulative errors that result in histopathological images with a complex mix of patterns and sub-patterns that only can be interpreted by an expert, even in cluttered biological circumstances. In addition, image capturing parameters such as environment illumination, exposure time, microscope magnification, etc., are a source of image variability. Therefore, the relevant visual pathological patterns highly change their visual appearance according to their spatial location, severity and co-occurrence with others biological structures. Figure 2-1 shows examples of some histopathological images of skin tissues with different annotations associated with acellular, cellular and

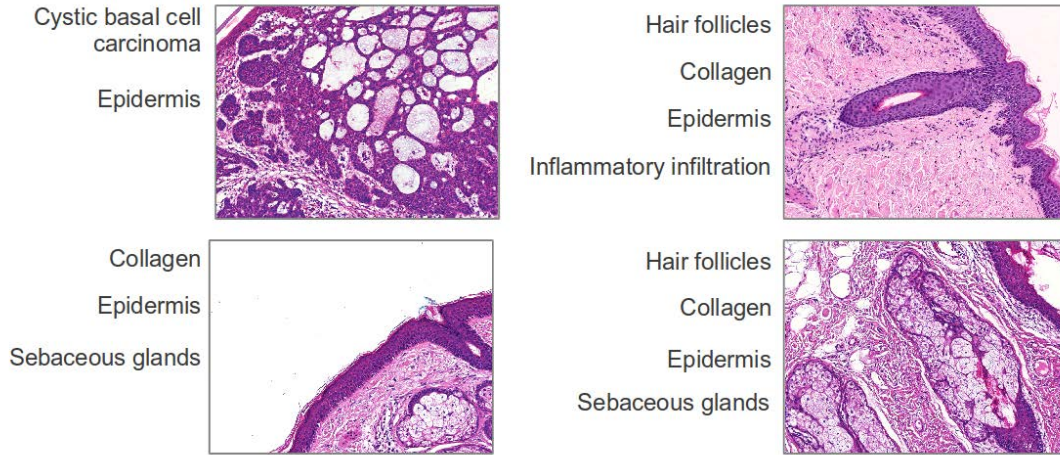


Figure 2-1: Example of histopathological images globally annotated with multiple annotations (multi-labeled images). These images correspond to the test data set used in this work and they have a resolution of  $1024 \times 768$  pixels. Histopathological annotations of morphological and architectural features such as epidermis, collagen and hair follicles appear in different images illustrating the high visual variability for the same annotation.

architectural features, illustrating the visual variability problem.

Commonly, bioimage analysis methods encompass two main components: a feature extraction and representation process that allows to properly describe the visual image content, which ideally should be robust to the visual variability problem of histopathological annotations, and an interpretable knowledge extraction process, capable of linking low-level visual patterns and high-level annotations. In this paper we propose a novel strategy for automatic annotation of histopathological images, which combines a part-based image representation (bag of features, BOF), a latent topic model (non-negative matrix factorization, NMF) and a probabilistic annotation strategy that allows to connect visual latent topics with high-level annotations. The proposed method provides both a robust automatic annotation method and a coarse location of them inside the images. The proposed method has a remarkable characteristic, it is exclusively trained with images that exhibit only one histopathological annotation. However the resulting annotation model is able to assign multiple histopathological annotations to full microscopical field of views. Therefore, it is not necessary to collect a representative training set that includes images that have different combinations of histopathological annotations. To the best of our knowledge, this is the first work that proposes an automatic annotation algorithm based on a part-based image representation and a

probabilistic latent topic model in histopathological images. The proposed approach was evaluated using a set of images of a skin cancer, known as *basal cell carcinoma*, which contains regions with ten different histopathological annotations, including acellular, cellular, and architectural features (i.e. *collagen*, *sebaceous glands*, *hair follicles*, *inflammatory infiltration*, *eccrine glands*, *epidermis*) and pathological lesions (i.e. *nodular basal cell carcinoma*, *morpheiform basal cell carcinoma*, *micro-nodular basal cell carcinoma*, *cystic basal cell carcinoma*).

The next sections are organized as follows: subsection 2.1.2 describes the proposed method based on BOF and NMF. Subsection 2.1.3 describes the basal-cell carcinoma data set and the performance measures used. Subsection 2.1.4 presents the evaluation results obtained for automatic annotation compared with a classical model of Support Vector Machines (SVM). Finally the conclusions are presented in Subsection 2.1.5.

## 2.1.2 Methodology

The proposed method for automatic annotation of histopathological images is depicted by the Figure 2-2. This approach comprises two main stages: i) training, and, ii) prediction. In the former stage, a probabilistic model that is able to automatically generate multiple annotations for new images (multi-label images) is generated from a set of images globally annotated with only one histopathological annotation (mono-label images). In this stage, a training set of images is represented by two matrices  $F$  and  $L$ , which codify the distribution of the visual information and the annotations of the images contained in the set, respectively. Note that,  $L$  will be a sparse matrix with 1 in the annotation assigned to each image and 0 in the other cases. The visual information is represented using a bag-of-features approach [109]. Therefore,  $F$  corresponds to a matrix of visual words versus images. To obtain an image representation using latent topics, the matrix  $F$  is factorized in two matrices ( $W$  and  $H$ ), using a NMF model that allows to find the probability distribution of visual latent topic models ( $H$ ) in the images. Finally, the visual latent topics are linked to the annotations distribution ( $L$ ) using a probabilistic model that generates a matrix ( $C$ ) with the latent representation of annotations. The prediction stage also starts with the bag of features representation of non-annotated images using the same visual codebook constructed in the training stage. A new image is projected to the latent topic space, given by  $W$ , to generate the vector  $H_{new}$ . Finally, the above vector ( $H_{new}$ ) and the latent representation of annotations ( $C$ ) are multiplied to obtain the vectors  $L_{new}$  that indicates the probability that the new image has each histopathological annotation. The new image is finally annotated with the corresponding histopathological annotation associated with one of the morphological features or pathological lesions with the highest probability by a



binarization process.

The details of the bag of features representation of images, latent topic model, and automatic-annotation process in training and prediction stages, using a probabilistic interpretation of non-negative matrix factorization, are introduced below.

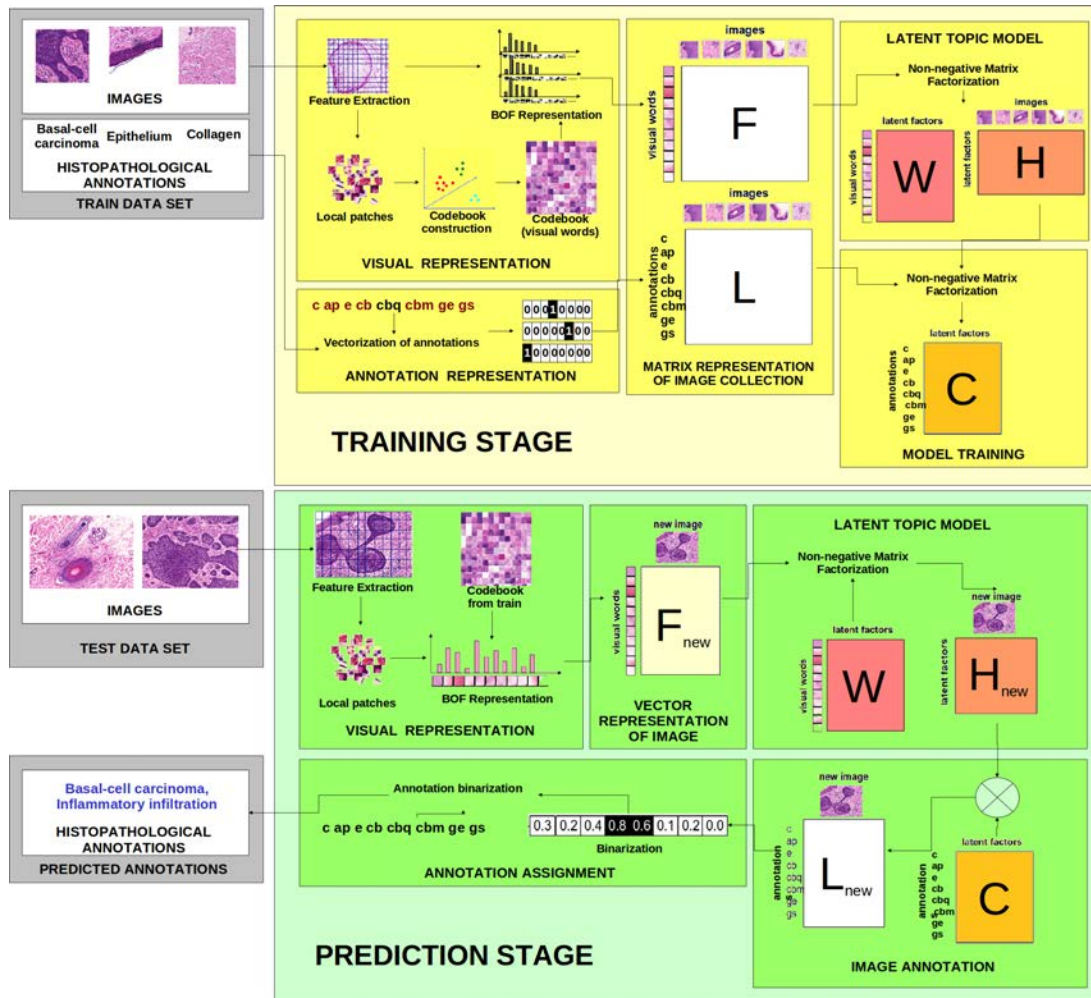


Figure 2-2: Overview of proposed method for automatic annotation of histopathological images based on non-negative matrix factorization.

### Bag of Features representation

The visual representation of histopathological images is obtained as a bag of features (BOF) [109]. A model inspired by the fact that the visual system perceives an object by integrating its constituent parts [17, 130]. Therefore, this representation is basically a histogram of small parts, called visual words, which are defined by a clustering analysis

of small patches extracted from an image collection. The general BOF representation approach comprises three main stages: feature detection and description, codebook or visual vocabulary construction, and BOF image representation [43].

Figure 2-3 depicts the BOF setup used in this work. The hypothesis underneath the proposed representation is that all biological structures are also represented by a probabilistic model that describe the distribution (histogram occurrence) of quantized small microstructures described by the visual patches. In the first step, the local feature detection consists in extracting small square patches that will be used for describing the whole visual image content. Herein, we extract these patches from a regular image partition of  $8 \times 8$  pixels without overlapping, which corresponds to the minimum resolution that a visual pattern require for covering biological structures such as cell nuclei. On the other hand, taking into account that from a pathologist point of view, visual identification of biological microstructures is based on the stain variations, we use the discrete cosine transform (DCT) coefficients of the RGB color components for describing each patch, because this local feature has been used to effectively describe this kind of variations in small regions [49, 89]. Local region descriptor results into a single feature vector of 192 dimensions by the concatenation of the three color component descriptors [37, 38, 50]. The second step, the codebook construction, is performed using a  $k$ -means clustering algorithm over a sample of patches from the training image set. The number of clusters,  $k$ , corresponds to the codebook size. The centroids found with this clustering are the visual words of the codebook and the visual representation of them can be obtained applying the inverse DCT. In this paper the codebook size was set to 700, which is a good value according to [38], where a systematic experimentation was performed on similar kind of images (histology and histopathology) using this visual features. Finally, each image is represented by a  $k$ -bin histogram. This is accomplished by associating the feature vector, describing each patch in the regular grid, to the closest visual word in the codebook. Then, the histogram is generated with each bin counting the number of patches in the image assigned to the corresponding visual word.

### Visual Latent Topic Analysis

The BOF representation of an image collection could be seen as a term-vs-document matrix,  $F \in \mathbb{R}^{n \times m}$ , where rows correspond to visual words and columns to images. Each element  $F_{ij}$  indicates the frequency of the  $i$ -th visual word in the  $j$ -th image. The goal of latent topic analysis is to find a set of latent factors that explain the visual content of each image as a mixture of different probability distributions of visual words. NMF is a well known matrix decomposition approach that approximates a matrix  $F \in \mathbb{R}^{n \times m}$  as a product of two simpler non-negative matrix factors  $W \in \mathbb{R}^{n \times k}$  and

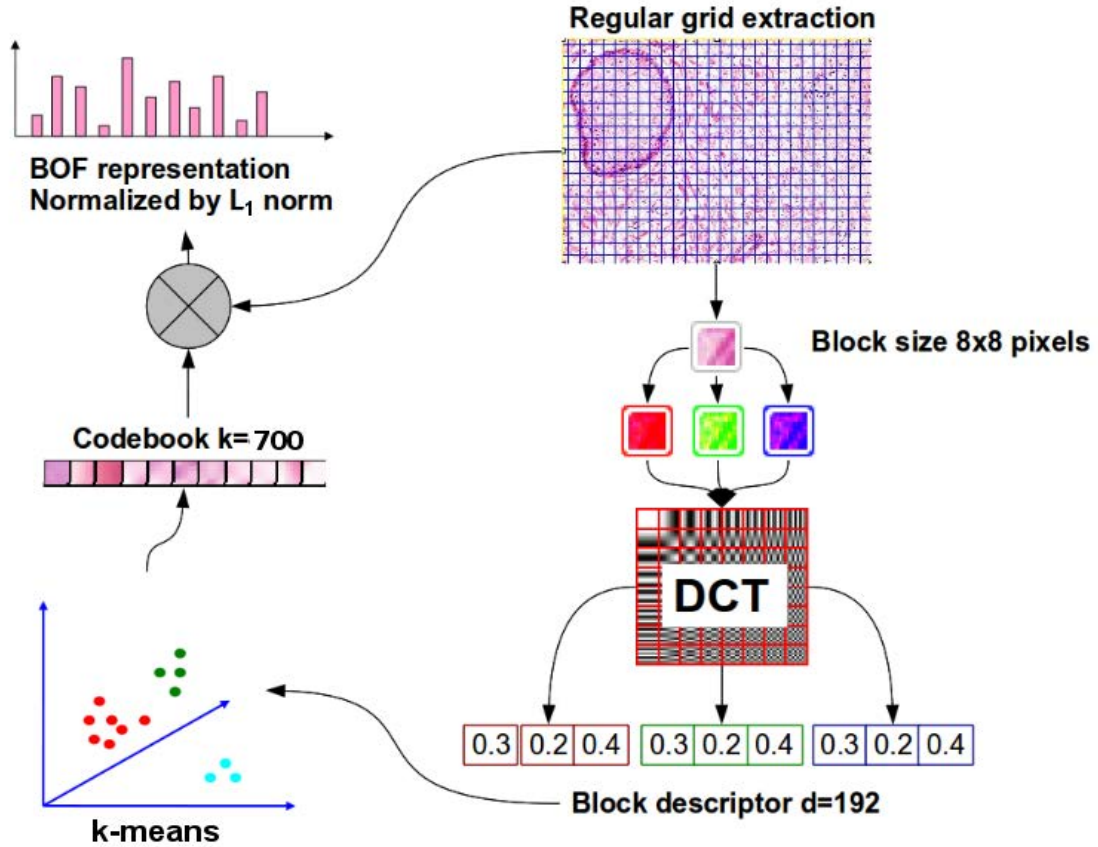


Figure 2-3: Bag of feature setup used for representing histopathology images. In this work the local features extraction is performed using regular grid extraction and each patch of  $8 \times 8$  pixels is represented by the first coefficients of a discrete cosine transform applied to each color component (RGB) independently, the visual codebook is built using  $k$ -means with  $k = 700$ , and finally each image is represented by a histogram of 700 bins normalized with  $L_1$  norm.

$H \in \mathbb{R}^{m \times k}$  as follows:

$$F = WH^T \quad (2-1)$$

with  $W$  containing a set of  $k$  latent factors that are linearly combined to represent the images in  $F$  using the coefficients in  $H$ .

The solution to NMF involves iterative optimization techniques using a cost function that describes the “closeness” of  $WH^T$  to  $F$ . Lee and Seung [104] proposed two different cost functions: Euclidean distance and Kullback Leibler (KL) divergence. In this work we use the last one because of its probabilistic interpretation [55]. The optimization problem based on KL divergence is defined as follows:

$$J = \underset{W, H}{\operatorname{argmin}} D(F || WH)_{KL} = \sum_{ij} F_{ij} \log \frac{F_{ij}}{[WH^T]_{ij}} - F_{ij} + [WH^T]_{ij} \quad (2-2)$$

An equivalence between NMF and probabilistic latent semantic indexing (PLSI) was reported by Ding et al. [55]. PLSI has a strong statistical foundation that models documents as a mixture of term probabilities conditioned on a latent random variable [80]. The parameters of the model are estimated by a likelihood maximization process based on expectation maximization algorithm. The mixture calculated by PLSI induces a factorization of the original term-document matrix:  $P(w_i, d_j) = \sum_{k=1}^r P(w_i | z_k) P(d_j | z_k) P(z_k)$ , if  $F$  is normalized according to  $F_{ij} \leftarrow F_{ij} / \sum_{ij} F_{ij}$ , it can be interpreted as the joint probability  $p(w_i, d_j) = F_{ij}$ . Ding et al. showed that the factorizations produced by NMF and PLSI are equivalent [55], with  $W$  containing the visual-word-latent-factor conditional probabilities,  $p(w_i | z_k)$ , and  $H$  the image-latent-factor joint probability,  $P(d_j | z_k) P(z_k) = P(d_j, z_k)$ .

In conclusion, NMF generates a model of the image collection that explains the occurrence of visual words in images by a mixture of probability distributions conditioned on a small set of latent factors. These latent factors can be interpreted as general visual patterns. Additionally, each latent factor can be associated to a cluster of images [54], the centroid of the cluster given by the columns of  $W$  and the assignment of images to clusters given by the rows of  $H$ , where the values can be interpreted as soft image-cluster membership functions.

### Probabilistic Annotation Model

Following the probabilistic model described in the previous subsection, the annotation task can be seen as the process of calculating the annotation-vs-image conditional

probabilities,  $p(l_c|d_{new})$ , where  $l_c$  is the  $c$ -th annotation and  $d_{new}$  corresponds to the new unannotated image. This is done by extending the latent topic model of the previous subsection with information from the annotations of the training images. This information is represented in an annotation-vs-image matrix,  $L \in \mathbb{R}^{c \times m}$ . The first step is to assign histopathological annotation to each one of the visual latent topics, i.e., to calculate the conditional probability  $p(l_c|z_k)$ . This is accomplished by applying NMF to the  $L$  matrix as follows:

$$L = CH^T \quad (2-3)$$

where  $H$  is the same matrix obtained from the visual latent topic factorization, which is kept fixed during the optimization process. After an appropriate normalization and according to the discussion of previous subsection,  $C$  contains the annotation-vs-latent-topic conditional probabilities  $p(l_c|z_k)$ . The second step is to project the new image to the visual latent space, this is done by applying NMF to solve:

$$F_{new} = WH_{new}^T \quad (2-4)$$

where  $F_{new}$  is the BOF representation of the new image,  $W$  is the same matrix obtained from the visual latent topic analysis and is kept fixed during the optimization process. After an appropriate normalization,  $H_{new}^T$  contains the joint probabilities  $p(d_{new}, z_t)$ . Finally, the conditional probability  $p(l_c|d_{new})$  is calculated using Bayes rule and law of total probability as follows:

$$p(l_c|d_{new}) = \frac{p(l_c, d_{new})}{\sum_c p(l_c, d_{new})} = \frac{p(l_c, d_{new})}{p(d_{new})} \quad (2-5)$$

where  $p(l_c, d_{new})$  is the factorized joint probability

$$p(l_c, d_{new}) = \sum_k p(l_c, d_{new}|z_k)p(z_k) = \sum_k p(l_c|z_k)p(d_{new}, z_k) \quad (2-6)$$

assuming that  $l_c$  and  $d_{new}$  are independent given  $z_k$ . It is easy to see that Equation 2-6 corresponds to the following matrix multiplication:

$$L_{new} = CH_{new}^T \quad (2-7)$$

According to the above discussion we propose a straightforward method for automatic annotation of images based on NMF (A2NMF) that consists in two stages (training and prediction) described in Algorithms 1 and 2.

---

**Algorithm 1** Training stage for automatic annotation of images using NMF.

---

1. Normalize  $F$  matrix to get joint probabilities of visual words and images.
  2. Normalize  $L$  matrix to get joint probabilities of histopathological annotations and images.
  3. Apply NMF with the visual information of the training data set (i.e.  $F$  visual-word-vs-image matrix) to get  $W$  and  $H$  matrices. Equation (2-1)
  4. Apply NMF with the annotation information of the training set (i.e.  $L$  annotation-vs-image matrix) fixing  $H$  matrix to get  $C$  matrix.
- 

---

**Algorithm 2** Prediction stage for automatic annotation of images using NMF.

---

1. Apply NMF with the visual information of new images (i.e.  $F_{new}$ ) fixing  $W$  matrix to get  $H_{new}$ . Equation (2-4)
  2. Multiply  $C$  and  $H_{new}$  matrices to get  $L_{new}$ . (Equation 2-7)
  3. Normalize  $L_{new}$  to get conditional probabilities  $p(l_c|d_{new})$   

$$L_{c,new} \leftarrow L_{c,new} / \sum_c L_{c,new}. \text{ Equation (2-5)}$$
  4. Binarize  $L_{new}$  assigning 1 if  $L_{c,new} > p(l_c)$  and 0 in otherwise
- 

### 2.1.3 Experimental design

#### Basal cell carcinoma data set

The proposed method was evaluated on a histopathological image data set, which was annotated by an expert, identifying the presence of architectural or morphological features, and pathological lesions inside each image. Images correspond to field of views with a 10X magnification, extracted from Hematoxylin-eosin ( $H\&E$ ) stained skin tissues diagnosed with different types of basal cell carcinoma. These images contain a particular richness in architectural and morphological features, i.e., characteristic arrangements of cells, surrounded by several combinations of epithelial and connective tissues, also found in many other pathologies [117].

The entire image set, composed of 655 digital images, was randomly divided into training (80%) and test (20%) sets. Square subimages that contained single histopathological annotations were manually cropped from the training image set. Although, there is no typical size for those annotations, because of their large intrinsic variabil-

ity, subimage size was estimated as an average value of a set of regular regions that the pathologist marked as containing a single histopathological annotation i.e. square subimages of  $300 \times 300$  pixels. A total of 1,466 training subimages were finally obtained, each containing a single annotation among the ten possibilities. On the other hand, the test set was composed of 138 images of  $1024 \times 768$  pixels, which, in general, are annotated as containing more than one histopathological annotation. Latter, images were globally annotated, i.e., the actual location of these annotations was not provided, which makes the task of automatic annotation even more challenging. The data set distribution by histopathological annotation is detailed in Table 2-1.

In order to reduce the visual image variability, a color normalization strategy, based on the transfer of the statistical properties of the stain contributions, was firstly applied [51]. Examples of some morphological features and a pathological lesion (*collagen*, *epidermis*, *hair follicles* and *cystic basal cell carcinoma*) are shown in Figure 2-4, in which the large appearance variability exposed by them can be appreciated. For example in the same figure Epidermis refers to outer layer of skin which comprises stratified squamous epithelium, i.e. several layers with different morphology of cells. However, typically a whole digital image of histopathology have one or more visual patterns associated with different morphological and architectural features of tissues like the images shown in Figure 2-1, which belong to the test image set. These images have in some located regions particular patterns, e.g. *epidermis* or *hair follicles*, whereas others are sparsely distributed without a well defined spatial location, e.g. *collagen*.

Table 2-1: Data set distribution per histopathological annotation for training and test.

Histopathological annotation	Train	Test	Total
Collagen (c)	337	70	407
Sebaceous glands (gs)	108	36	144
Hair follicles (ap)	106	33	139
Inflammatory infiltration (i)	135	90	225
Eccrine glands (ge)	108	22	130
Epidermis (ep)	144	39	183
Nodular Basal cell carcinoma (cb)	208	33	241
Morpheiform basal cell carcinoma (cbm)	132	14	146
Micro-nodular basal cell carcinoma (cbn)	83	9	92
Cystic basal cell carcinoma (cbq)	105	9	114
Total	1,466	138	1,604



Figure 2-4: Examples of training images with the corresponding histopathological annotations. These images have a resolution of  $300 \times 300$  pixels and exhibit only one annotation per image.

### Performance evaluation

The performance of the proposed automatic annotation method was evaluated using standard measures such as precision, recall, accuracy and f-measure which are defined as follows:

$$Precision = \frac{tp}{tp + fp}, Recall = \frac{tp}{tp + fn}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}, F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where  $tp$  is the number of correctly predicted annotations,  $fp$  is the number of wrong predicted annotations,  $tn$  is the number of correctly omitted annotations, and  $fn$  is the number of missed annotations.

Note that these measures are not evaluated independently by class. This is a better way to evaluate the performance in an automatic annotation task where images simultaneously exhibit multiple annotations.

As baseline, a state-of-the-art supervised annotation method based on support vector machines (SVM) was used. We train a one-vs-all SVM model with an RBF kernel for each class, the best parameters were chosen using a 10-fold cross-validation over the



training data set. As well the proposed approach, the SVM model uses the same BOF image representation for visual content of the images.

The performance of the proposed and baseline annotation methods was evaluated in both training and testing data sets. When evaluating the performance in the training data set, 20% of the training images are withheld during training and later used to evaluate the generalization performance. The purpose of this two-way evaluation was to contrast the performance of the annotation methods in two scenarios: a simple mono-label annotation task, corresponding to annotate images with the same characteristics as that ones used for training, and the original complex multi-label annotation task.

### 2.1.4 Results and discussion

An important parameter for a latent-topic model is the size of latent space dimension, i.e the number of latent topics required for representing the collection visual content. Therefore, the effect of varying the number of latent topics was assessed on the mono-label annotation task. Figure **2-5** shows the average performance of the proposed method in the training data set against the number of latent topics (dimension of the latent space). With a small number of latent dimensions, the annotation model has a high recall, but with low precision, accuracy and f-measure. This indicates that the annotation model tends to assign a high number of annotations per image. The situation improves with a higher number of latent dimensions, and all the measures steadily increase beyond 32 dimensions.

These results suggest selecting a  $k$  value as big as possible. Lee and Seung in [103] suggested a number of latent topics  $k < nm/(n + m)$ . The reason is that beyond of this value the number of parameters in the factorization, the maximum number of dimensions in the latent space according to the rule  $(n + m)k$ , will be greater than the number of values in the original matrix,  $nm$ . We decided to use this limit for the experiments, taking into account that  $n$  is the codebook size and  $m$  is the number of images in training stage. This gives a number of latent topics  $k = 438$  for the mono-label scenario and  $k = 473$  for the multi-label scenario.

The proposed approach was compared against a state-of-the-art SVM model in both mono-label and multi-label scenarios. Table **2-2** shows the average value for each performance measure, for both scenarios on the respective test set. The results show that the SVM model performs better on the mono-label annotation tasks. The reason could be because the test images are similar to those used in the training stage, i.e. small images containing a unique annotation. However, when test images contain more than one histopathological annotation, the proposed approach takes advantage of the intermediate representation in the latent semantic space, and outperforms the results

reported by the SVM learning model.

The results suggest that the proposed model is doing a better work characterizing the high visual variability of the different histopathological annotations. A supervised learning model, such as SVM, requires a representative set of training images that exhibit combination of morphological and architectural features of tissues similar to the ones expected in the test set. In contrast, the proposed method initially characterizes the visual variability of the training data set in an unsupervised fashion. Annotations of the training data set are used in a later step to build a probabilistic annotation model that connects latent visual topics with histopathological annotations.

One important characteristic of the proposed method for automatic annotation is its interpretability. Whereas SVM is one of the most powerful models for supervised learning, the generated classifiers are not easily interpretable. The improved interpretability of the proposed method is due to the fact that it is possible to map back the generated labels to particular regions of the image by each morphological and architectural feature. This is accomplished by assigning a histopathological annotation posterior probability to each small image patch. Figure **2-6** illustrate the concept mapping strategy: a test image is shown in Figure **2-6a** and the corresponding probabilities maps of its patches for each of the ten histopathological annotations are shown in Figure **2-6b-k**. Each of these maps have in the top the real binary membership value of the histopathological annotation ( $v$ ), the posterior probability of the predicted annotation given the image by the proposed method in Equation 2-5 ( $p$ ), and the binary classification of image with the corresponding histopathological annotation according to the step 4 of Algorithm 2 ( $b$ ).

These results are relevant in the biomedical context because the high-variability of architectural and morphological features in healthy and pathological tissues is a common phenomena. In general, it is very difficult to have enough examples of each possible structural arrangement of the morphological features for training a supervised learning model such as the SVM algorithm. This scenario is also a more realistic in biomedical image domain, where regions of interest in the image, which cover an example of biological structures, are commonly annotated by the presence or absence of a given set of histopathological annotations whereas computer-aided diagnosis or image retrieval systems require the annotation of full images.

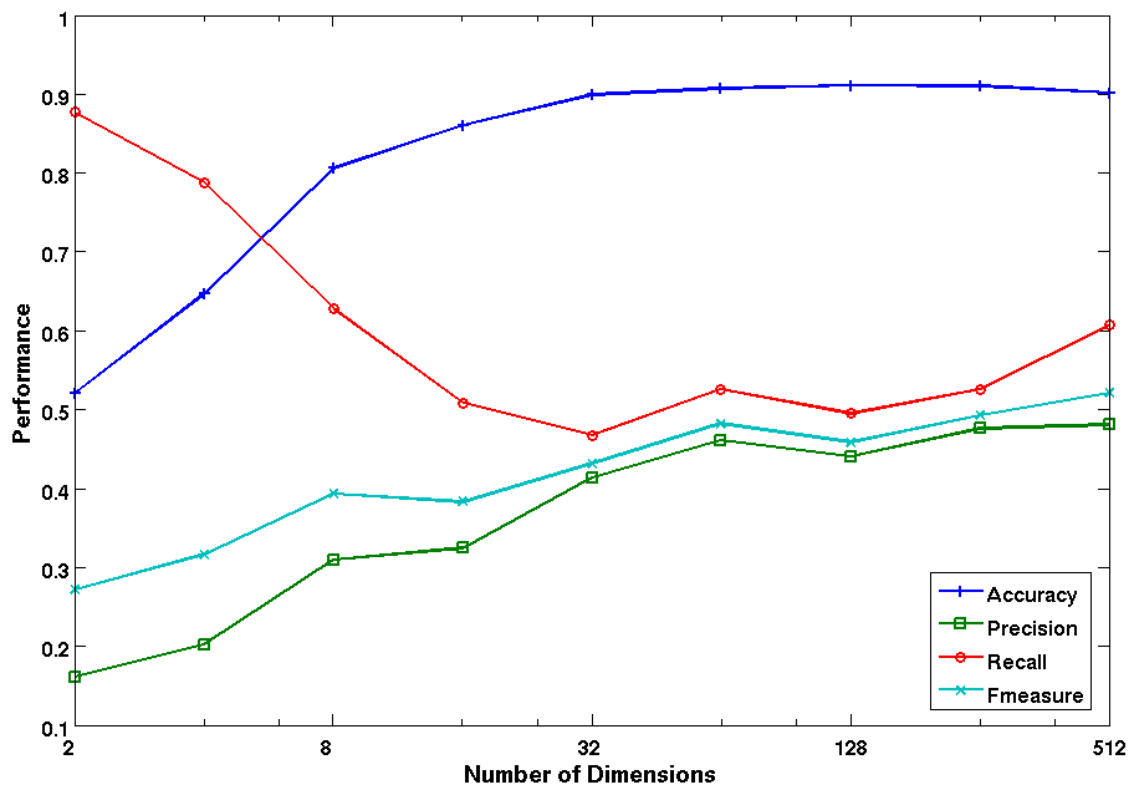


Figure 2-5: Performance evaluation on training mono-label images by each number of dimensions in the latent space.

Table 2-2: Average in automatic annotation performance in both experiments with standard performance measures, Accuracy (Acc), Precision (Pr), Recall (Rc) and F-measure (F)

	Mono-label images				Multi-label images			
Method	Acc	Pr	Rc	F	Acc	Pr	Rc	F
SVM-RBF	<b>0.96</b>	<b>0.84</b>	<b>0.69</b>	<b>0.76</b>	0.70	0.26	0.10	0.11
A2NMF	0.92	0.67	0.46	0.51	<b>0.76</b>	<b>0.5</b>	<b>0.74</b>	<b>0.55</b>

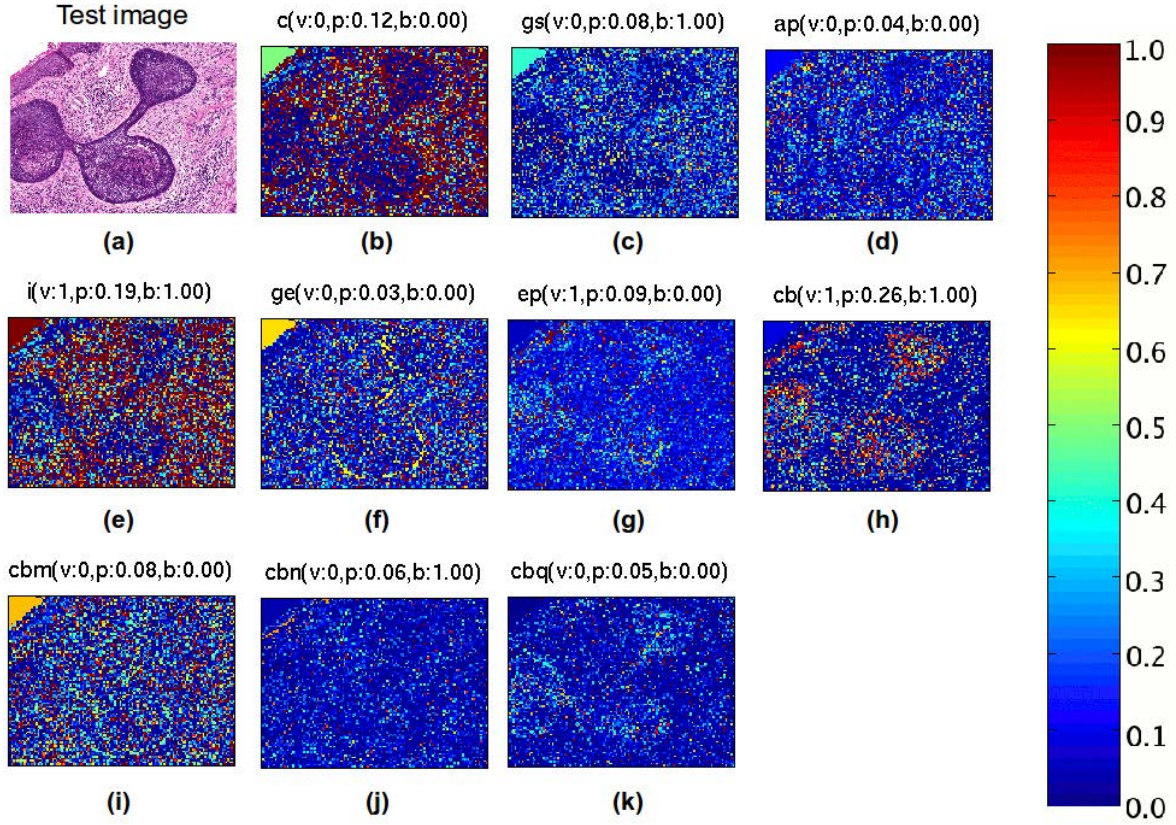


Figure 2-6: Example of an image from the test data set automatically annotated by the proposed method. The original multi-label image (a) is showed with the salient maps of the patches inside the image according with each one of the ten histopathological annotations: *collagen* (b), *sebaceous glands* (c), *hair follicles* (d), *inflammatory infiltration* (e), *eccrine glands* (f), *epidermis* (g), *nodular basal cell carcinoma* (h), *morpheiform basal cell carcinoma* (i), *micro-nodular basal cell carcinoma* (j), *cystic basal cell carcinoma* (k)), on the top of each salient image is the real membership of the class (v), the conditional probability estimated by the proposed method (p) and the final concept binarization value (b).

### 2.1.5 Conclusions

We presented a novel method for histopathological images annotation with probabilistic support for prediction and spatial location of morphological and architectural features in healthy and pathological tissues. The method was evaluated in a challenging scenario where training images corresponded to small subimages exhibiting only one histopathological annotation, although test images included multiple annotations. The proposed

method exhibited an improved performance when compared to a state-of-the-art supervised annotation method. The distinctive characteristic of the proposed method is that it builds an enhanced representation of the visual image collection content in an unsupervised fashion finding latent visual topics, which encode high-level visual patterns.

Histopathological images are particularly challenging to analyze because of their high variability and complex visual structure. The results reported here suggest that latent semantic characterization of the visual structure is a viable alternative to build competitive annotation models for histopathological images.

## 2.2 Anaplastic Tumor Differentiation and Visual Interpretation in Medulloblastoma

A method for automatic analysis and interpretation of histopathology images is presented. The method uses a representation of the image data set based on bag of features histograms built from visual dictionary of Haar-based patches and a novel visual latent semantic strategy for characterizing the visual content of a set of images. One important contribution of the method is the provision of an interpretability layer, which is able to explain a particular classification by visually mapping the most important visual patterns associated with such classification. The method was evaluated on a challenging problem involving automated discrimination of medulloblastoma tumors based on image derived attributes from whole slide images as anaplastic or non-anaplastic. The data set comprised 10 labeled histopathological patient studies, 5 for anaplastic and 5 for non-anaplastic, where 750 square images cropped randomly from cancerous region from whole slide per study. The experimental results show that the new method is competitive in terms of classification accuracy achieving 0.87 in average. *The complete content of this section has been published as a research article in the proceedings of 15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'2012) (see [42]).*

### 2.2.1 Introduction

This paper presents a new method, ViSAI, for automatic analysis and interpretation of histopathological images. The method comprises three main stages: learning of an image representation based on bag of features (BOF), characterization of the rich visual variety of a histopathological image collection using visual latent topic analysis, and connection of visual patterns with the semantics of the problem using a probabilistic classification model. The learnt probabilistic model is applied to new images, and the

class posterior probability is used to determine the corresponding class. The method is applied to the classification of a type of brain cancer called medulloblastoma, which is one of the most common types of malignant brain tumors [141]. In adults, the disease is rare whereas in children the incidence amounts to a 25% of all pediatric brain tumors. Tumor classification of medulloblastoma is currently performed by microscopical examination and no quantitative image analysis and classification tools are so far available for this task. Different histologic types of medulloblastoma have different prognosis. The differential diagnosis is a hard task and tends to be qualitative. Determine the subtypes of medulloblastoma are difficult to establish and subject to inter-observer variability because of the similarity between the two basic histologic subclasses: anaplastic and non-anaplastic and their similarity with a long list of differential diagnoses. The anaplastic medulloblastoma have worse prognosis and this is mostly characterized by the presence of large, irregular cells that lack organization and in some cases attempt to wrap around each other. The therapeutical management changes radically depending on the subtype of medulloblastoma so that histopathological diagnosis is useful in determining the potential outcome of the disease. Hence computerized and quantitative image analysis tools are useful in this kind of problem for better estimation of medulloblastoma subtype allowing potentially make better prognostic decisions.

Recent investigations [114] have pointed out the importance of provide computerized and automatic image analysis tools as support to diagnosis and prognosis for different types of cancer. Recent work in histologic image analysis has explored the use of automated methods for breast cancer and prostate cancer diagnosis and grading [44,95] using other images modalities. Galaro et al. [69] classified anaplastic and non-anaplastic medulloblastoma subtypes, using a BOF of Haar wavelet coefficients as local pixel-based descriptors. The authors of [69] reported on average a classification accuracy of 0.80. The above works have essentially developed tools to improve the quality of diagnosis in terms of objective support and some level of quantification. Bag of features approach is an image content representation commonly used in computer vision area which comprises three main stages: local feature extraction, dictionary construction, and image representation. This approach have been successful adapted and applied in histology images previously in image analysis and classification tasks [38,40]. On the other hand, latent topic analysis is an approach widely used in text document collection analysis to find the semantic topics related with these documents. The representative techniques are latent semantic analysis (LSA), pLSA [80] and latent Dirichlet analysis (LDA) [18]. Both pLSA and LDA suppose a generative model for documents. The main assumptions here are: first, the image content of image could be represented by an appropriate set of visual words from dictionary learnt from whole image collection represented by a good visual word representation, and second, the

large visual variability in the image collection is generated from a relatively small set of visual latent factors. Under our analysis, visual latent factors correspond to high-level patterns that mix sets of visual words that co-occur with high frequency in the collection. Semantics is then linked to visual latent factors coding the relationship between the visual appearance and particular classes.

The new method in this paper addresses the problem of automated classification of histological slides. The method provides higher accuracy coupled with an interpretation of classification results for the expert in terms of semantic domain, rather than being a black-box approach. In particular, the main contributions of the present work are:

- A strategy to characterize the rich visual content of histology images combining an image representation, based on BOF, and a texture local descriptor for image patches based on a Haar wavelet transform.
- A visual latent-topic analysis model for finding higher-level visual patterns that combine multiple visual words.
- A method that is able, not only to globally classify a virtual slide, but also provides interpretability determining the most important visual patterns associated with such classification, and identifying their location in the original image.

### 2.2.2 Methodology

ViSAI method performs an implicit semantic identification of the visual patterns that characterize each class (anaplastic and non-anaplastic). The strategy is built upon an image representation, based on BOF, and a visual latent semantics analysis, based on probabilistic latent semantic analysis (pLSA). The overall scheme is depicted in Figure 2-7.

#### Image Representation

**Step 1: Local features extraction.** square patches of 50 pixels, that is the minimum spatial resolution that covers a cellular unit (nucleus and cytoplasm), are extracted using a dense grid strategy with an overlap of 80%. Each patch is represented by two different local descriptors, raw-blocks (block) and Haar wavelet coefficients (haar). The raw-block descriptor corresponds to luminance values of pixels in the patch, whereas the Haar descriptor corresponds to the filter responses of two scales of a Haar-wavelet-based wavelet transform. Luminance differences are reduced by applying, before feature extraction, a mean-zero-variance-one normalization.



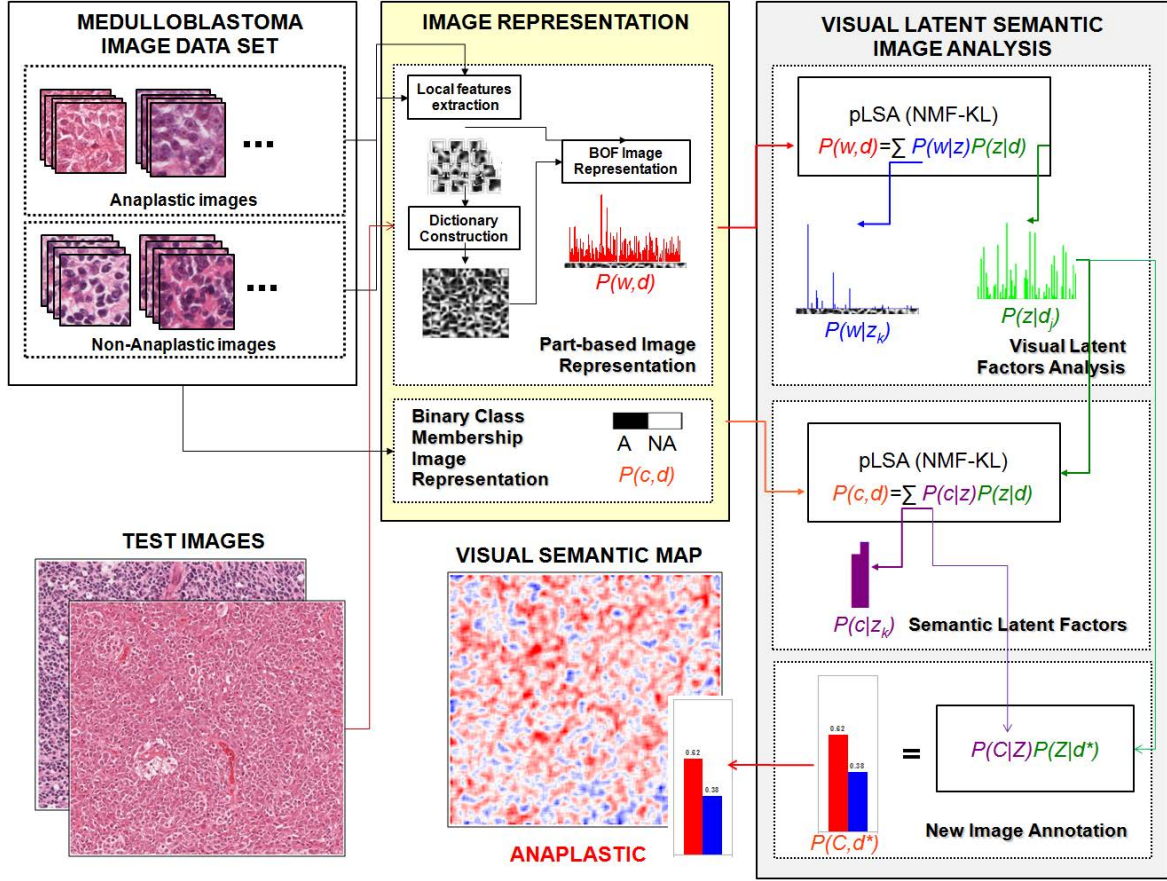


Figure 2-7: Overall scheme for the visual latent semantic analysis for automatic classification and interpretation method.

**Step 2: Dictionary construction.** the dictionary is built by applying a conventional k-means clustering strategy to a sample of image patches from the training set. The number of clusters,  $k$ , corresponds to the dictionary size, and each centroid corresponds to a visual concept. An important difference with a previous texon-based approach [69] is that our dictionary is constructed from images randomly selected from both classes, regardless of whether it belongs to the class anaplastic or not, by which the visual concepts are mixed up within a very heterogeneous dictionary.

**Step 3: Novel Image representation.** each image is then represented by a  $k$ -bin histogram, capturing the frequency of each visual word in the image given the local descriptor and the visual dictionary. A collection of images can be represented then by a matrix,  $X$ , where each column corresponds to an image and each row to a visual word. If each column is normalized (L1 normalization), each position in the matrix,  $X_{i,j}$ , corresponds to the conditional probability of finding a visual word  $w_i$  in an image



$d_j$ , i.e.,  $p(w_i|d_j)$ .

### Visual Latent Semantic Image Analysis

**Step 1: Visual latent factors analysis.** The new method uses an approach similar to pLSA that assumes that the presence of different terms (visual words) in a set of documents (images) can be explained by the presence of a reduced set of hidden variables called latent factors. Specifically, the conditional probability  $P(w_i|d_j)$  can be represented as:

$$P(w_i|d_j) \approx \sum_k P(w_i|z_k)P(z_k|d_j) \quad (2-8)$$

where  $P(w_i|z_k)$  is the conditional probability of visual word  $w_i$  given latent factor  $z_k$  and  $P(z_k|d_j)$  is the conditional probability of a latent factor  $z_k$  given the image  $d_j$ . The latent factors can be found by solving an optimization problem that looks for  $P(W|Z)$  and  $P(Z|D)$  that minimizes the Kullback-Leibler divergence between the left and right sides of Equation 2-8 [55]. In our case, we solve the optimization problem by modeling it as a matrix factorization problem:  $P(W|D) = P(W|Z)P(Z|D)$ , where  $P(W|D) = X$  is the histogram representation of the image collection discussed in Subsection 2.2.2,  $P(W|Z)$  contains the latent factors represented in terms of visual words, and  $P(Z|D)$  contains the representation of the images in the collection, in terms of latent factors.

**Step 2: Semantic latent factors.** Given a set of annotated training images, each image,  $d_i$ , is associated with one of two classes: anaplastic ( $C_1 = A$ ) or non-anaplastic ( $C_2 = NA$ ). If, for instance, the image is anaplastic then  $P(C_1|d_i) = 1$  and  $P(C_2|d_i) = 0$ . Following the same reasoning as in previous subsection this probability can be expressed in terms of the latent factors as follows:

$$P(C_i|d_j) \cong \sum_k P(C_i|z_k)p(z_k|d_j) \quad (2-9)$$

The probability  $P(C_i|z_k)$  effectively connects each latent factor with the semantics of the problem represented in terms of the biological concept associated with each class. These probabilities can be found applying a matrix factorization algorithm [55] that fixes  $P(Z|D)$  from the previous step and looks for  $P(C|Z)$  such that:  $P(C|D) = P(C|Z)P(Z|D)$ .

**Step 3: New image classification.** A new image,  $d^*$ , is first represented using the strategy discussed in Subsection 2.2.2. This produces a normalized histogram  $P(W|d^*)$ . The image is then represented in terms of latent factors ( $z_k$ ) finding  $P(Z|d^*)$  such that:  $P(W|d^*) \cong P(W|Z)P(Z|d^*)$ , where  $P(W|Z)$  was previously found in step 1 solving Equation 2-8. Then, the posterior class probability of the image  $P(C|d^*)$  is calculated using:  $P(C|d^*) \cong P(C|Z)P(Z|d^*)$ , where  $P(C|Z)$  was previously calculated in step 2 solving Equation 2-9 and obtained  $P(Z|d^*)$ . Finally, the class assigned to the new image is the one with the maximum class posterior probability.

### 2.2.3 Experimental design

The dataset comprises 10 labeled histopathological cases from St. Jude Children’s Research Hospital, which 5 are anaplastic and 5 are non-anaplastic. Each slide is a whole virtual slide of  $80000 \times 80000$  pixels with one or more cancerous regions with a large tumoral variability, manually annotated by a neuro-pathologist. For every slide, 750 individual images of  $200 \times 200$  pixels non-overlapping were extracted uniformly at random from these cancerous regions, resulting in a database of 7500 different images: half of them anaplastic. The local feature extraction is carried out as was described in Subsubsection 2.2.2 for two local descriptors *block* and *haar*. The dictionary size was tested with different sizes, 10, 20, 40, 80, 160 and 320 visual words. Finally the number of latent factors was fixed to the dictionary size since this amounts to a maximum number of elemental building blocks (visual words) to be represented by visual semantic concepts (latent factors). The probabilistic analysis described in Subsubsection 2.2.2 provides a probability value associated with the semantic importance of each of the initial visual features so that these can be grouped together (into latent factors) by similar levels of relevance per class (Equations 2-8 and 2-9).

In this sense, three different experiments were proposed to evaluate each parameter of the proposed method. First, determine the visual words appearance from dictionaries and impact of dictionary size for each local descriptor (block or haar) evaluation average accuracy in test data set over multiple trials of cross-validation. Second, evaluate the classification performance of the proposed image representation (Subsubsection 2.2.2) for each local descriptor using the proposed method (Subsubsection 2.2.2) compared with a k-NN classifier. Third, a visual mapping of high-level concepts to identify spatial regions associated by each one (anaplastic and non-anaplastic).

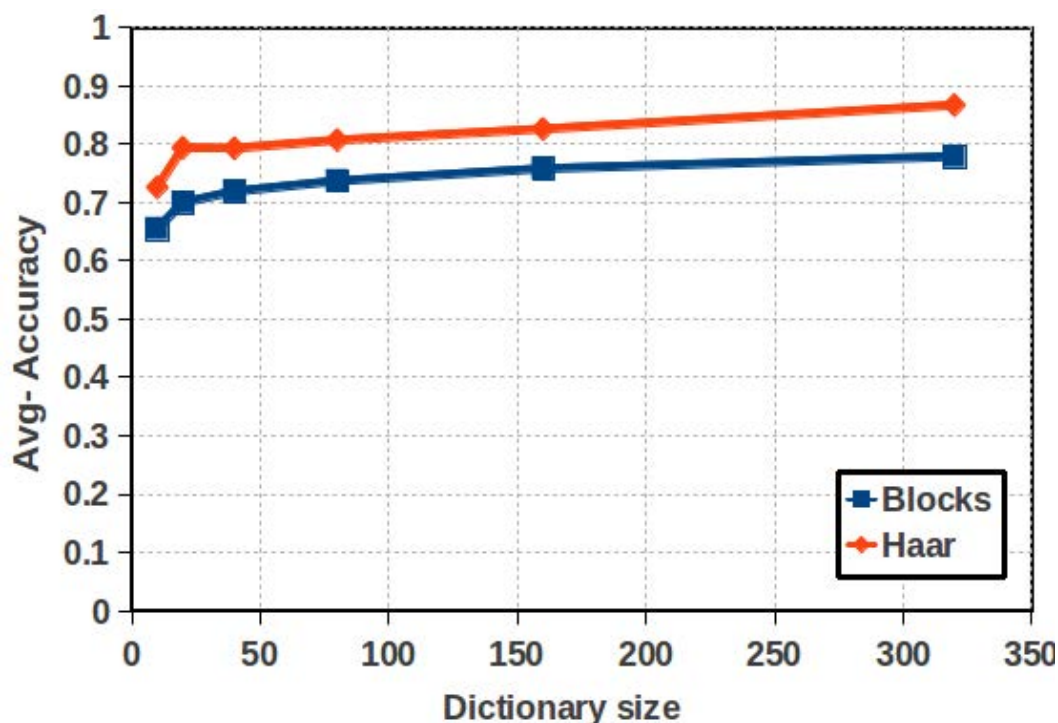


Figure 2-8: Dictionary size vs average accuracy for both local features (block and haar).

## 2.2.4 Results and discussion

### Visual dictionary construction

Figure 2-8 shows the impact of the dictionary size on the validation average accuracy. Overall, haar-based dictionary outperforms the raw-block representation by about 7%. For both representations the best performances were obtained with the largest dictionary size obtaining 0.77 (blocks) and 0.86 (haar) in average accuracy. Visual words can be related to a particular class calculating the posterior class probability. Table 2-3 shows the 20 visual words with highest posterior class probability for both classes, for the two types of representation. In all the cases the visual dictionary is composed basically of smooth samples of nuclei and cells in different orientations and arrangements, a result that matches perfectly with a very basic principle in pathology analysis which introduces the cell as the basic unit of information. Also some differences between the visual words associated to each class can be observed, for both (block and haar), non-anaplastic patterns are more homogeneous. This is consequent with the biological definition of anaplasia, where cells are undifferentiated and nuclei shapes and sizes have high variability.

Table 2-3: A sample of the visual dictionaries obtained using different local features. Second column shows the 20 visual words with highest probability by class (A:anaplastic, NA:non-anaplastic) for each kind of local feature in a dictionary size of 160.

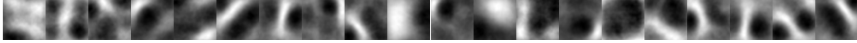
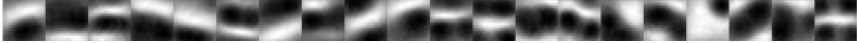


Feature	20 visual words with highest probability by class																			
block	A																			
	NA																			
haar	A																			
	NA																			

Table 2-4: Classification performance in terms of accuracy, specificity and sensitivity for dictionary size of 320.

	ViSAI			k-NN		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
block	0.78	<b>0.89</b>	0.67	0.80	0.79	0.81
haar	<b>0.87</b>	0.86	<b>0.87</b>	0.84	0.88	0.79

### Classification performance

For each of multiple trials of cross-validation, we used 4 anaplastic and 4 non-anaplastic slides for training the visual semantic model and 1 anaplastic and 1 non-anaplastic slides for testing. For comparison, the  $k$ -NN classifier was employed. An optimal value for  $k$ , 10, was found by cross validation such as was suggested in [69]. Different dictionary sizes and patch representations were evaluated to determine the best configuration for this classification task in terms of classifier accuracy, specificity, and sensitivity. The results are presented in Table 2-4, which shows the performance obtained with each image representation strategy and classification algorithm. Clearly the haar-based representation outperforms the block-based representation, independent of classifier choice. The new classification method is competitive with respect to the  $k$ -NN, a classical classifier used in [69]. The improvement obtained by our method using haar-based representation was 3.6% in accuracy and 10.1% in specificity.

### Visual pattern mapping and interpretation

The methodology presented in this paper allows for new unknown images to determine the spatial probability maps for each class using conditional probabilities. Figure

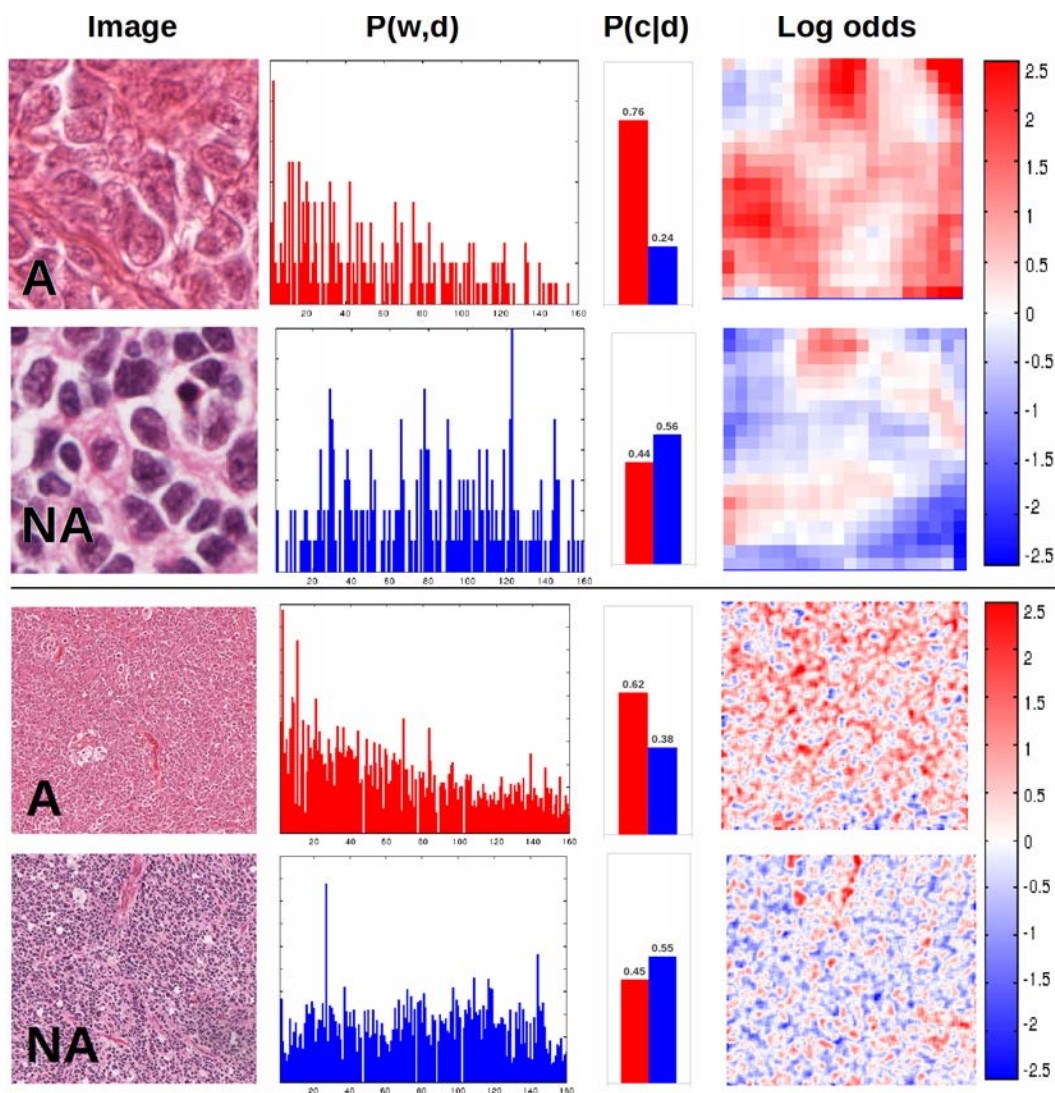


Figure 2-9: Visual semantic maps for test images from a cross validation trial. Images in the first column were correctly classified (A-anaplastic and NA-non-anaplastic). See Subsection 2.2.4 for description.

**2-9** illustrates the semantic visual maps generated for two unknown test images in a particular trial of cross validation. Rows one and two show images of the same size used in training ( $200 \times 200$ ), whereas rows three and four show a larger field of view ( $2000 \times 2000$ ). The center column shows the corresponding BOF histogram representation for each image (i.e. occurrence frequency of each visual word of the dictionary in the image). Bin (visual words) has been sorted according to the posterior class probabilities: visual words with a higher class probability of being anaplastic are at the left, and visual words with a higher class probability for non-anaplastic are at the right. The third column shows the posterior class probability. The fourth column shows a map that indicates how likely a particular region of the image is to be related to one of the semantic classes. This is depicted by the log-odds of the posterior class probabilities to emphasize the differences. A positive value (*red*) indicates a higher probability for class anaplastic, a negative value (*blue*) indicates a higher probability for class non-anaplastic. An advantage of the BOF image representation is that new method is able to scale towards larger image size as shown in the rows three and four of Figure **2-9**, where the correct prediction of classes and spatial location of semantic regions is more challenging.

### 2.2.5 Conclusions

The new method ViSAI was successfully applied to the challenging problem of automatic discrimination between two subtypes of medulloblastoma using computer derived image features extracted from whole slides. The goal of our method was not just to improve classification accuracy, but to provide interpretable results using latent semantic analysis and BOF image representation. The method provides an interpretation layer that helps out the pathologist to determine the type of patterns present in the sample under examination with a potential improvement of diagnostic significance. This attribute is particularly useful in problems relating to stratification of the disease where the distinction between disease sub-classes might reside in very subtle visual cues. The experimental results are promising and indicate that visual latent semantic analysis has potential as a tool for analyzing the complex visual patterns exhibited by histopathology images.

## **Part II**

# **Feature and Representation Learning**

## 3 Unsupervised histopathology image representation learning

*The two main digital pathology tasks are tumor detection and tumor grading. Pathological analysis and diagnosis focus on tumor regions. The tumor aggressiveness is determined by tumor differentiation and grading according to particular morphological and arrangement properties of tumoral cells. This chapter presents methods for unsupervised representation learning applied to the analysis of histopathology images of skin (see into Section 3.1) and brain cancer (see into Section 3.2) for tumor detection and grading. The following sections correspond to papers presenting the details of the methods and applications.*

### 3.1 Tumor Classification of Basal Cell Carcinoma

This paper presents and evaluates a deep learning architecture for automated basal cell carcinoma cancer detection that integrates (1) image representation learning, (2) image classification and (3) result interpretability. A novel characteristic of this approach is that it extends the deep learning architecture to also include an interpretable layer that highlights the visual patterns that contribute to discriminate between cancerous and normal tissues patterns, working akin to a digital staining which spotlights image regions important for diagnostic decisions. Experimental evaluation was performed on set of 1,417 images from 308 regions of interest of skin histopathology slides, where the presence of absence of basal cell carcinoma needs to be determined. Different image representation strategies, including bag of features (BOF), canonical (discrete cosine transform (DCT) and Haar-based wavelet transform (Haar)) and proposed learned-from-data representations, were evaluated for comparison. Experimental results show that the representation learned from a large histology image data set has the best overall performance (89.4% in F-measure and 91.4% in balanced accuracy), which represents an improvement of around 7% over canonical representations and 3% over the best equivalent BOF representation. *The complete content of this section has been published in the 16th International Conference on Medical Image Computing and Computer Assisted Intervention - MICCAI 2013 (see [33]), and extended*



versions had been published in *IX International Seminar on Medical Information Processing and Analysis - SIPAIM 2013* and *Journal of Artificial Intelligence in Medicine* (see [4, 5]).

### 3.1.1 Introduction

This paper presents a unified method for histopathology image representation learning, visual analysis interpretation, and automatic classification of skin histopathology images as either having basal cell carcinoma or not. The novel approach is inspired by ideas from image feature representation learning and deep learning [16] and yields a deep learning architecture that combines an autoencoder learning layer, a convolutional layer, and a softmax classifier for cancer detection and visual analysis interpretation. Deep learning (DL) architectures are formed by the composition of multiple linear and non-linear transformations of the data, with the goal of yielding more abstract – and ultimately more useful – representations [16]. These methods have recently become popular since they have shown outstanding performance in different computer vision and pattern recognition tasks [16, 93, 98]. DL architectures are an evolution of multilayer neural networks (NN), involving different design and training strategies to make them competitive. These strategies include spatial invariance, hierarchical feature learning and scalability [16]. An interesting characteristic of this approach is that feature extraction is also considered as a part of the learning process, i.e., the first layers of DL architectures are able to find an appropriate representation of input images in terms of low-level visual building blocks that can be learnt.

This work addresses the challenging problem of histopathology image analysis and in particular the detection of basal-cell carcinoma (BCC), the most common malignant skin cancer and may cause significant tissue damage, destruction and, in some cases, disfigurement. Unlike natural scene images, where typical automated analysis tasks are related to object detection and segmentation of connected regions that share a common visual appearance (e.g. color, shape or texture), histopathology images reveal a complex mixture of visual patterns. These patterns are related to high variability of biological structures associated with different morphology tissue architecture that typically tend to significantly differ in normal and diseased tissue. Another source of visual variability is the acquisition process itself, going from a 3D organ biopsy to a 2D sample (histopathological slide). This process involves different stages: sampling, cutting, fixing, embedding, staining and digitalization, each one contributing inherent artifacts [40]. Figure 3-1 shows histopathology image samples stained with hematoxylin-eosin (H&E) from cancerous and non-cancerous tissue samples. These images illustrate the high intra-class visual variability in BCC diagnosis, caused by the

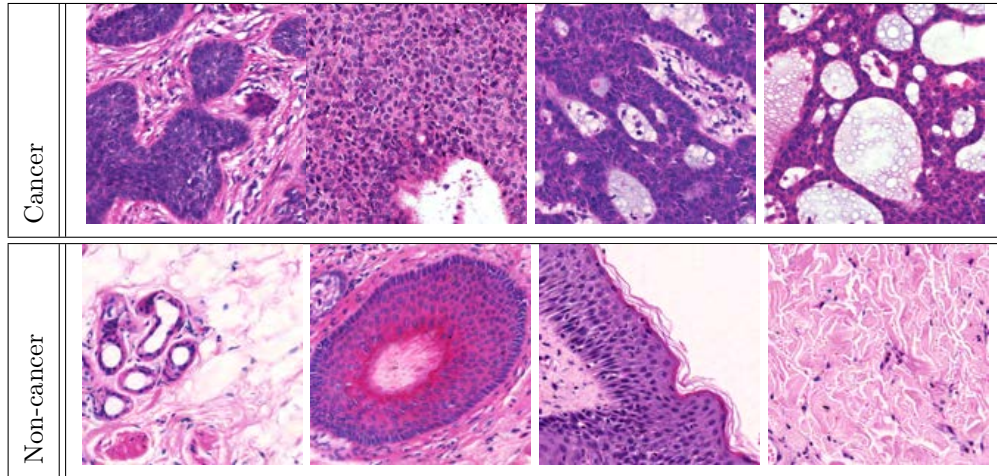


Figure **3-1**: Example of BCC histopathology images (both cancer and non-cancer) stained with H&E at 10X.

presence (or absence) of different morphological and architectural structures, healthy (eccrine glands, hair follicles, epithelium, collagen, sebaceous glands) or pathological (morpheiform, nodular and cystic change).

There is an extensive literature in automatic histopathology image analysis where different strategies for image representation have been tried: discrete cosine transform (DCT), wavelet coefficients, Gabor descriptors, and graph representations among others [106]. In all cases, the goal is to capture the visual features that better characterize the important biological structures related to the particular problem been addressed. This means that some representations may work better than others depending on the particular problem. On account of recent advances in computer vision [93, 98], there is an encouraging evidence (mostly for natural scene images) that learned representations (induced from data) may have a better performance than canonical, predefined image feature representations. To the best of our knowledge, this is the first attempt to evaluate learned-from-data image features in BCC histopathology image analysis using a DL architecture. A related approach is a bag of features (BOF) representation, which attempts to learn a set of visual code words from training data and uses them to represent images. However, this representation strategy still needs the definition of a local feature descriptor in advance (e.g. raw-block, SIFT histogram, DCT coefficients). Previous research has suggested that the particular local feature descriptor choice has an important effect on BOF performance [38, 42].

This paper presents a convolutional auto-encoder DL architecture for histopathology image classification as a tool to support BCC diagnosis. The DL architecture is enhanced by an interpretation layer that highlights the image regions that most contribute

in the discrimination of healthy tissue from cancer. The main novel contributions of this work are:

- A BCC histopathological image analysis method, which integrates, in a unified DL model, the following functionalities: image feature representation learning, image classification, and prediction interpretability.
- An evaluation of learned-from-data image representations in BCC histopathology image analysis, which shows that this approach could produce improved classification performance while enhancing model interpretability.
- A novel strategy to exploit the information in the intermediate representation layers of the DL architecture to produce visual interpretable predictions. In that sense this method is analogous to a digital stain which attempts to identify image regions that are most relevant for making diagnostic decisions.

While there has been some previous related work in the use of DL architectures for automatic segmentation and classification in histopathology images for breast cancer cell detection, unlike our approach, the methods in [121, 133] do not focus on image representation learning. Hence these methods do not analyze the learned features and do not explore their potential for visual prediction interpretation. Prediction interpretability is not an important issue when analyzing natural images, so it has not been typically studied in classical computer vision literature. However, in the context of systems for performing predictions and decision support there is a need to explain and identify those visual patterns which are relevant for prediction. While some approaches have been for visual prediction interpretation [38, 40, 42, 51], these approaches have not used a DL architecture and in all of them the interpretation ability is provided by an additional stage, subsequent to image classification. By contrast, in our approach, visual interpretability is tightly integrated with the classification process.

### 3.1.2 Methodology

The new method for histopathology image representation learning, BCC cancer classification and interpretability of the results of the predictor, is based on a multilayer neural network (NN) architecture depicted in Figure 3-2. The different stages or modules of the framework, corresponding to different layers of the NN are described as follows:

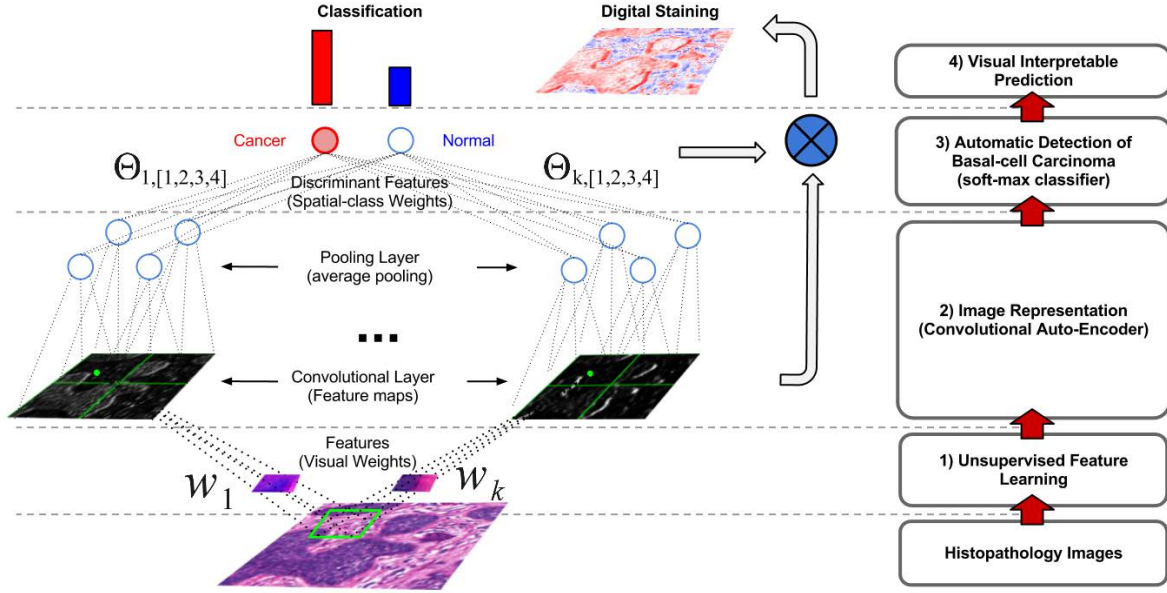


Figure 3-2: Convolutional auto-encoder neural network architecture for histopathology image representation learning, automatic cancer detection and visually interpretable prediction results analogous to a *digital stain* identifying image regions that are most relevant for diagnostic decisions.

### Step 1. Unsupervised feature learning via autoencoders:

Training images are divided into small patches ( $8 \times 8$  pixels), which are used to train an autoencoder NN [16] with  $k$  hidden neurons. This produces a set of weight vectors  $W = \{w_1, \dots, w_k\}$ , which can be interpreted as image features learned by the autoencoder NN. The autoencoder looks for an output as similar as possible to the input [16]. This autoencoder learns features by minimizing an overall cost function with a sparsity constraint to learn compressed representations of the images defined by:

$$J_{sparse}(W) = J(W) + \beta \sum_{j=1}^k KL(\rho || \hat{\rho}_j), \quad (3-1)$$

where  $J(W)$  is the typical cost function used to train a neural network,  $\beta$  controls the weight of sparsity penalty term.  $KL(\rho || \hat{\rho}_j)$  corresponds to Kullback–Leibler divergence between  $\rho$ , desired sparsity parameter, and  $\hat{\rho}_j$ , average activation of hidden unit  $j$  (averaged over the training set).

### Step 2. Image representation via convolution and pooling:

Any feature  $w_k$  can act as a filter, by applying a convolution of the filter with each image to build a feature map. The set of feature maps form the convolutional layer. Thus, a particular input image is represented by a set of  $k$  features maps, each showing how well a given pattern  $w_i$  spatially matches the image. This process effectively increases the size of the internal representation ( $\approx k \times$  the size of the original representation) of the image. The next layer acts in the opposite direction by summarizing complete regions of each feature map. This is accomplished by neurons that calculate the average (pool function) of a set of contiguous pixels (pool dimension). The combination of convolution and pooling provide both translation invariance feature detection and a compact image representation for the classifier layer.

### Step 3. Automatic detection of BCC via softmax classifier:

A softmax classifier, which is a generalization of a logistic regression classifier [93], takes as input the condensed feature maps of the pooling layer. The classifier is trained by minimizing the following cost function:

$$J(\Theta) = -\frac{1}{m} \left[ \sum_{i=1}^m y^{(i)} \log h_{\Theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\Theta}(x^{(i)})) \right], \quad (3-2)$$

where  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$  is the corresponding training set of  $m$  images, where the  $i$ -th training image is composed of  $y^{(i)}$  class membership and  $x^{(i)}$  image representation obtained from the output of the pooling layer, and  $\Theta$  is a weight vector dimension  $k \times n$  (where  $n$  is the pool dimension). The output neuron has a sigmoid activation function, which produces a value between 0 and 1 that can be interpreted as the probability of the input image being cancerous.

### Step 4. Visual interpretable prediction via weighted feature maps:

The softmax classifier weights ( $\Theta$ ) indicate how important a particular feature is in discriminating between cancer and non-cancer images. A weight  $\Theta_j$  (associated<sup>1</sup> to a feature  $w_k$ ) with a high positive value indicates that the corresponding feature is associated with cancer images, in the same way a large negative value indicates that the feature is associated with normal images. This fact is exploited by the new method to build a *digitally stained* version of the image, one where cancer (or non-cancer)

---

<sup>1</sup>In general, a feature has as many weights associated with it as the pool dimension. When the pool dimension is larger than one, the average weight is used.

related features are present. The process is as follows: each feature map in the convolutional layer is multiplied by the corresponding weight ( $\Theta$ ) in the softmax layer, all the weighted feature maps are combined into an integrated feature map. A sigmoid function is applied to each position of the resulting map and finally the map is visualized by applying a colormap that assigns a blue color to values close to 0 (non-cancer) and a red color to values close to 1 (cancer).

### 3.1.3 Experimental design

#### Histopathology basal cell carcinoma dataset description (BCC dataset):

The BCC dataset comprises 1417 image patches of  $300 \times 300$  pixels extracted from 308 images of  $1024 \times 768$  pixels, each image is related to an independent ROI on a slide biopsy. Each image is in RGB and corresponds to field of views with a 10X magnification and stained with H&E [40, 51]. These images were manually annotated by a pathologist, indicating the presence (or absence) of BCC and other architectural and morphological features (collagen, epidermis, sebaceous glands, eccrine glands, hair follicles and inflammatory infiltration). The Figure 3-1 shows different examples of these images.

#### Learned image feature representations:

The main focus of the experimental evaluation was to compare the image representations learned from the histopathology data, generated by the DL-based proposed method, against two standard canonical features (discrete cosine transform (DCT) and Haar-based wavelet transform (Haar)). Since the focus of this experimentation was the evaluation of different image representations, the same classification method was used in all the cases (steps 2 to 4 of the new architecture, Figure 3-2). Also a comparison against BOF image representation was included with the same number of features and patch size employing the same local feature descriptors (DCT and Haar) based on previous work [40, 42]. In the case of canonical image representations, the feature weights  $\{w_1, \dots, w_k\}$  were replaced by the basis vectors that define either DCT or Haar. In addition to the image features learned from the BCC dataset, two other sets of image features were learned from different data sets: a histology data set composed of healthy histological tissue images (HistologyDS<sup>2</sup> [38]) and a natural scene image data set commonly used in computer vision research (STL-10 dataset<sup>3</sup>). In or-

<sup>2</sup>Available in: <http://www.informed.unal.edu.co/histologyDS/>

<sup>3</sup>Available in: <http://www.stanford.edu/~acoates/stl10/>

der to choose an appropriate parameter configuration for our method, an exhaustive parameter exploration was performed. The parameters explored were: image scales (50%,20%), number of features (400, 800), pool dimension (7, 13, 26, 35, 47, 71, 143) and pooling function (average or sum). The best performing set of parameters were a image scale of 50%, 400 features, a pool dimension of 71 with average pooling function, and all the reported results correspond to this configuration. A patch size of  $8 \times 8$  pixels was used since this was ascertained to be the minimum sized region for covering a nucleus or cell based off the recommendations from previous related work using the same datasets [40, 42].

### **Cancer detection performance:**

Once the best parameter configuration for the DL architecture is selected, each representation was qualitatively evaluated for comparison using a stratified 5-fold cross-validation strategy on a classification task (discriminating cancer from non-cancerous images) in the BCC dataset. The performance measures employed were accuracy (Acc), precision (Pr), recall/sensitivity (Rc/Sen), specificity (Spc), f-measure (F1) and balanced accuracy (BAC).

## **3.1.4 Results and discussion**

### **Learned representations vs canonical representations:**

Figure 3-3 shows a set of 200 image features learned from three different data sets, (a) BCC, (b) HistologyDS and (c) STL-10, along with the set of features corresponding to and obtained from the DCT representation, (d). In all the cases, features were sorted according to their frequency when representing the BCC data set. As expected features learned from BCC and HistologyDS images better capture visual patterns related to dyes, edges of large nuclei in different orientations and perhaps most interestingly small dots related to common/healthy nuclear patterns that do not appear in the other feature sets. Features learned from the natural image dataset (STL-10), also capture visual patterns such as edges, colors, and texture but are less specific. The features associated with DCT representation are even less specific and only capture general color and texture patterns.

### **Automatic cancer detection performance:**

Table 3-1 presents the classification performance results in terms of accuracy, precision, recall, specificity, f-measure and balanced accuracy (BAC). The results show a

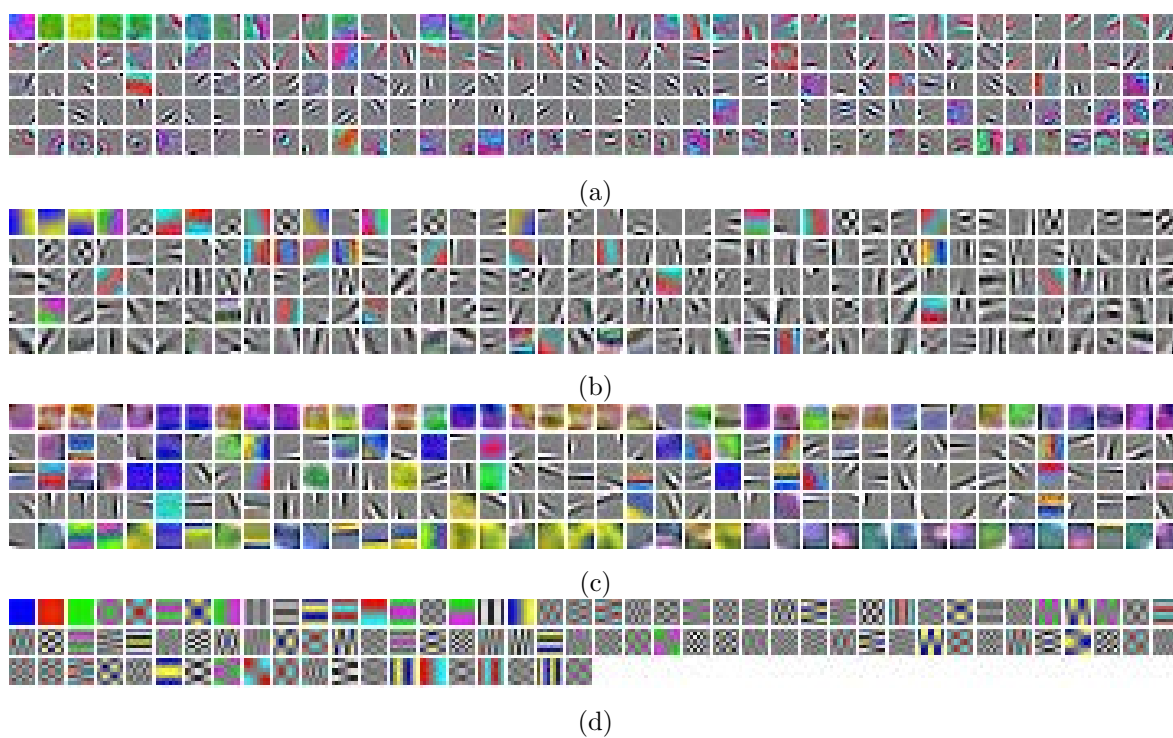


Figure **3-3**: Comparison of learned features (a.k.a. dictionaries or basis) by autoencoders from: a) BCC (histopathology), b) HistologyDS (healthy tissues) and c) STL-10 (natural scenes) datasets, and d) DCT basis.



Table **3-1**: Classification performance for each learned features from different image datasets using the new DL based method, canonical features, along with different configurations of BOF. The best results are in bold typeface.

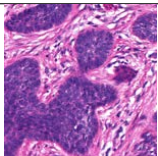
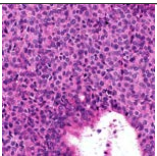
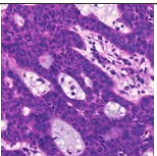
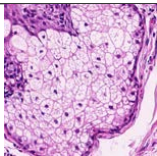
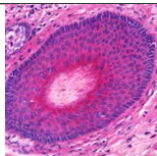
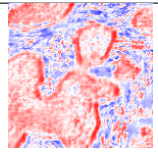
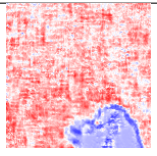
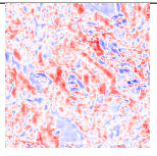
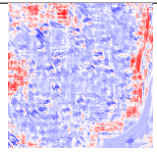
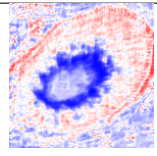
	Acc	Pr	Rc/Sen	Spc	F1	BAC
(DL) BCC	0.90+/-0.03	0.87+/-0.03	0.86+/-0.04	0.92+/-0.02	0.87+/-0.03	0.89+/-0.03
(DL) HistologyDS	<b>0.92+/-0.03</b>	<b>0.90+/-0.04</b>	<b>0.88+/-0.03</b>	<b>0.94+/-0.03</b>	<b>0.89+/-0.03</b>	<b>0.91+/-0.02</b>
(DL) STL-10	0.90+/-0.02	0.87+/-0.04	0.86+/-0.02	0.92+/-0.03	0.86+/-0.02	0.89+/-0.02
(DL) DCT	0.86+/-0.03	0.82+/-0.04	0.79+/-0.05	0.90+/-0.03	0.80+/-0.03	0.84+/-0.03
(DL) Haar	0.84+/-0.03	0.78+/-0.03	0.78+/-0.06	0.87+/-0.04	0.78+/-0.02	0.82+/-0.03
(BOF) Haar-400	0.79+/-0.02	0.79+/-0.05	0.68+/-0.06	0.86+/-0.04	0.70+/-0.03	0.77+/-0.02
(BOF) GrayDCT-400	0.88+/-0.02	0.88+/-0.03	0.83+/-0.04	0.90+/-0.01	0.83+/-0.02	0.87+/-0.02
(BOF) ColorDCT-400	0.89+/-0.02	0.89+/-0.02	0.85+/-0.03	0.91+/-0.01	0.85+/-0.02	0.88+/-0.02

clear advantage of learned features over canonical and BOF representations. A  $t$ -test showed that differences among DL models with learned features was not significant ( $p > 0.05$ ) and that DL models were significantly better ( $p < 0.05$ ) than canonical features and BOF representations. The fact that features learned from histology images produced the best results for histopathology image classification is an interesting result, suggesting that the proposed approach is learning important features to describe general visual pattern present in different histopathological images. This is consistent with the findings in [138], which had shown that the strategy of learning features from other large datasets (known as self-taught learning) may produce successful results .

### Digital staining results:

Table **3-2** illustrates some examples of the last and most important stage of the new DL method– digital staining. The table rows show from top to bottom: the real image class, the input images, the class predicted by the model, the probability associated with the prediction, and the digital stained image. The digitally stained version of the input image highlights regions associated to both cancer (red stained) and non-cancer (blue stained) regions. These results were analyzed by a pathologist, who suggested that our method appeared to be identifying cell proliferation of large-dark nuclei. A caveat however is that this feature also appears to manifest in healthy structures where the epidermis or glands are present. Nonetheless, this enhanced image represents an important addition to support diagnosis since it allows the pathologist to understand why the automated classifier is suggesting a particular class.

Table 3-2: Outputs produced by the system for different cancer and non-cancer input images. The table rows show from top to bottom: the real image class, the input image, the class predicted by the model and the probability associated to the prediction, and the digital stained image (red stain indicates cancer regions, blue stain indicates normal regions).

True class	Cancer	Cancer	Cancer	Non-cancer	Non-cancer
Input image					
Pred/Prob	Cancer (0.82)	Cancer (0.96)	Cancer (0.79)	Non-cancer (0.27)	Non-cancer (0.03)
Digital staining					

### 3.1.5 Conclusions

We presented a novel unified approach for learning image representations, visual interpretation and automatic BCC cancer detection from routine H&E histopathology images. Our approach demonstrates that a learned representation is better than a canonical predefined representation. This representation could be learned from images associated with the particular diagnostic problem or even from other image datasets. The paper also presented a natural extension of a DL architecture to do digital staining of the input images. The inclusion of an interpretability layer for a better understanding of the prediction produced by the automated image classifier.

## 3.2 Tumor Differentiation of Medulloblastoma Brain Cancer

Learning data representations directly from the data itself is an approach that has shown great success in different pattern recognition problems, outperforming state-of-the-art feature extraction schemes for different tasks in computer vision, speech recognition and natural language processing. Representation learning applies unsupervised and supervised machine learning methods to large amounts of data to find

building-blocks that better represent the information in it. Digitized histopathology images represents a very good testbed for representation learning since it involves large amounts of high complex, visual data. This paper presents a comparative evaluation of different supervised and unsupervised representation learning architectures to specifically address open questions on what type of learning architectures (deep or shallow), type of learning (unsupervised or supervised) is optimal. In this paper we limit ourselves to addressing these questions in the context of distinguishing between anaplastic and non-anaplastic medulloblastomas from routine haematoxylin and eosin stained images. The unsupervised approaches evaluated were sparse autoencoders and topographic reconstruct independent component analysis, and the supervised approach was convolutional neural networks. Experimental results show that shallow architectures with more neurons are better than deeper architectures without taking into account local space invariances and that topographic constraints provide useful invariant features in scale and rotations for efficient tumor differentiation. *The complete content of this section has been published as a research article in the **Tenth International Symposium on Medical Information Processing and Analysis - SIPAIM 2014** (see [31]).*

### 3.2.1 Introduction

Recently, there has been a surge of interest in representation learning [16], owing to their improved performance over state of the art machine learning approaches. These approaches, which implicitly involve finding optimally class separable data representations, differ from classical supervised learning approaches in that they do not require prior feature extraction or classifier training.

Recently there has been a great deal of interest in developing pattern recognition schemes in order to identify prognostic patterns for disease from digitized images of histopathology slides [113]. Most of these approaches have focused on a traditional pipeline of feature extraction, feature selection, and classification to identify patterns and regions of interest from very large and complex histopathology images [68]. However this may not always be the optimal approach, particularly in scenarios where it may not always be obvious what the pattern of interest is. Most pattern recognition approaches in digital pathology therefore involve some form of supervised learning (e.g. tumor and mitosis detection [28,36,116]) and relatively few approaches are geared towards unsupervised learning [5,33].

However, it is still not clear as to what the best methods for digital pathology tasks are and the most desirable properties for these models in this domain. For natural scene images, several papers have investigated the application of representation learn-

ing [16,93,101]. These comparative studies have focused on the use of more data, deeper architectures (more hierarchical layers), and role of pooling layers in providing spatial invariance for object detection. In histopathology images there are still many open questions regarding what the best learning approach is, supervised or unsupervised. Additional questions linger as to which is the architecture that works best, deeper or shallow, and what assumptions or properties of models are appropriate, spatial, scale, and/or rotation invariance. Like most machine learning methods, representation learning is mainly categorized as either unsupervised and supervised. Among unsupervised representation learning methods there are several types of autoencoders (AE), such as Sparse AE, Denoise AE, Reconstructed Independent Component Analysis (RICA) or Topographic ICA (TICA). Whereas supervised representation learning methods are mostly configured as Convolutional Neural Networks (CNN). In this paper we present a comparative evaluation of the most popular representation learning methods applied to histopathology images in the context of the challenging problem of distinguishing between anaplastic and non-anaplastic medulloblastoma.

The primary novel contributions of this work include: 1) the first comparative evaluation of representation learning approaches, supervised and unsupervised, for problems in digital pathology, 2) new findings about how topographic properties of unsupervised learning methods yields improved detection results, 3) the first successful application of representation learning methods to medulloblastoma tumor differentiation.

The rest of this paper is organized as follows: Section 2 describes supervised and unsupervised representation learning methods; Section 3 presents the medulloblastoma dataset used for validation, the experimental setup, the performance measures and the corresponding evaluation results; Section 4 summarizes the novel contributions, whereas in Section 5 we present our concluding remarks and directions for future work.

### 3.2.2 Methodology

#### **Supervised feature learning approach: Convolutional Neural Networks**

A CNN [102] is a multilayer neural network with an architecture that comprises one or more consecutive convolutional and pooling layers followed by a final full-connected layer. The convolution layer applies a 2D convolution of the input feature maps and a convolution kernel. The pooling layer applies a L2 pooling function over a spatial windows without overlapping (pooling kernel) per each output feature map, L2 pooling, in particular, allows the learning of invariant features. The output of the pooling layer is the input of a fully-connected layer which mixes them into a feature vector. The outputs of the full-connected layer are two neurons (anaplastic, non-anaplastic)

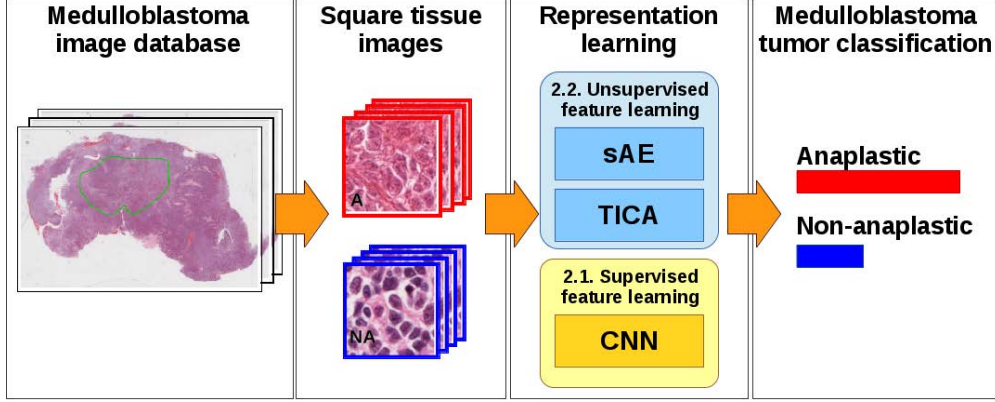


Figure 3-4: Overview of representation learning framework for medulloblastoma differentiation between anaplastic and non-anaplastic tumors by comparing unsupervised and supervised feature learning approaches.

activated by a logistic regression model. For this work we evaluate different CNN architectures with only one convolutional and pooling layer (shallow) or two convolutional and pooling layers (deeper), varying also the number of features (neurons) per layer and learned patch sizes (feature kernels). The whole CNN model is trained using Stochastic Gradient Descent to minimize the loss function:

$$J_{CNN}(\Theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\Theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\Theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^N \Theta_{ij}^2, \quad (3-3)$$

where  $m$  is the number of samples in training dataset,  $k$  is the number of classes,  $x^{(i)}$  is the output of full-connected layer corresponding to  $i$ -th sample,  $y^{(i)}$  is the class label of  $i$ -th sample,  $1\{.\}$  is the indicator variable,  $\Theta$  is the parameters of logistic model. The second term corresponds to weight decay which penalizes large values of the parameters. During the training process, weight parameters in all layers are updated by the backpropagation algorithm.

### Unsupervised feature learning approaches: Sparse Autoencoders and TICA

In contrast to supervised feature learning methods, like CNN, unsupervised feature learning approaches learn a set of feature detectors (feature kernels) that can explain better the content directly from the data without taking into account the class labels associated. Similarly to supervised approaches, the input data is represented by linear or non-linear transformations of learned features.

There are different approaches to perform unsupervised feature learning (UFL). Among the most popular are variants of auto-encoding UFL, which are methods that learn an

encoding function, built from learned features, able to reconstruct its input. Here we evaluate two of them: Sparse Autoencoders and Topographic Reconstruct Independent Component Analysis.

**Sparse Autoencoders (sAE):** One of the inherent properties that is desirable for an efficient and compact representation of data is sparseness. It means to represent the input data by fewer elements of learned features as possible. The objective function to learn this representation for sAE is given by:

$$J_{sparse}(\Theta) = \frac{1}{2} \sum_{i=1}^m \|r_{\Theta}(x^{(i)}) - x^{(i)}\|_2^2 + \beta \sum_{j=1}^n KL(\rho || \hat{\rho}_j) + \frac{\gamma}{2} (\|\mathbf{W}\|_F^2 + \|\mathbf{W}'\|_F^2), \quad (3-4)$$

where  $x^{(i)} \in \mathbb{R}^d$  is the  $i$ -th sample in the training set  $\mathbf{X}$ ,  $\beta \in \mathbb{R}$  is the hyperparameter that controls the trade-off between regularization and reconstruction, and  $KL(\rho || \hat{\rho}_j) = \rho \log\left(\frac{\rho}{\hat{\rho}_j}\right) + (1 - \rho) \log\left(\frac{1-\rho}{1-\hat{\rho}_j}\right)$  is the Kullback-Leibler divergence between two Bernoulli distributions with means  $\rho \in \mathbb{R}$ , the desired sparsity parameter percentage, and  $\hat{\rho}_j \in \mathbb{R}$ , the average activation of the  $j$ -th feature detector.  $\Theta = \{\mathbf{W}, \mathbf{W}', b, c\}$  is the set of parameters,  $\mathbf{W} \in \mathbb{R}^{n \times d}$  and  $\mathbf{W}' \in \mathbb{R}^{d \times n}$  are the encoder and decoder weight matrices, and  $b \in \mathbb{R}^n$  and  $c \in \mathbb{R}^d$  are encoder and decoder bias vectors. Finally, the third term regularizes magnitudes of the weights through the Frobenius norm ( $\|\cdot\|_F$ ), with the hyperparameter  $\gamma \in \mathbb{R}$  controlling the importance of the term in the objective function.

**Topographic Reconstruct Independent Component Analysis (TICA):** Topographic models seek to organize learned features such that similar activations are close together, while different ones are set apart. This arrangement follows the visual cortex model where cells have a specific spatial organization and response of neurons change in a systematic way [5]. Particularly, TICA builds a square matrix to organize feature detectors in  $l$  groups such that adjacent feature detectors activate in a similar proportion to the same stimulus. TICA cost function is given by:

$$J_{TICA}(\mathbf{W}) = \frac{\lambda}{m} \sum_{i=1}^m \|\mathbf{W}^T \mathbf{W} x^{(i)} - x^{(i)}\|_2^2 + \sum_{i=1}^m \sum_{k=1}^l \sqrt{\mathbf{H}_k (\mathbf{W} x^{(i)})^2 + \epsilon}, \quad (3-5)$$

where  $l$  is the number of desired groups in the topography,  $\mathbf{H} \in \{0, 1\}^{l \times n}$  is the topographic organization with  $\mathbf{H}_k^{(j)} = 1$  if the  $j$ -th feature detector belongs to the  $k$ -th group, 0 otherwise. This model sets  $\mathbf{H}$  fixed and learns  $\mathbf{W}$ . TICA is unconstrained and it can be treated with efficient optimization solvers like L-BFGS. TICA calculates two types of features: basic features,  $f_j(x) = \mathbf{W}_j x$ , and invariant features  $f_j^*(x) = \sqrt{\mathbf{H} (\mathbf{W} x)^2}$ . Invariant features group several basic features based on their adjacency in the topographic map.

### 3.2.3 Experimental design

#### Medulloblastoma histopathology cancer dataset

The database is composed by 10 pathology slide cases (patients) stained with hematoxylin and eosin (H&E) diagnosticated as Medulloblastoma Cancer from St. Jude Children’s Research Hospital (5 anaplastic and 5 non-anaplastic). Each slide is a whole-slide image of  $80,000 \times 80,000$  pixels with one or more cancerous regions manually annotated by a neuropathologist. The final dataset was built by uniform random sampling without overlapping of 750 square regions ( $200 \times 200$  pixels) per slide, resulting into 7,500 square regions (3,750 anaplastic and 3,750 non-anaplastic). For this work, all square regions were converted to grayscale images and normalized by mean zero and unit variance, because color is not discriminative enough and luminance variation support.

#### Experimental setup

In order to evaluate the performance of each representation learning approach, we applied the same experimental design described in [42, 69]. For comparison, baseline methods are the ones described in those previous works. Hence, the evaluation was done by multiple trials of cross-validation. For each trial, a subset of 4 anaplastic and 4 non-anaplastic slides randomly selected were used for training whereas the remaining slides, 1 anaplastic and 1 non-anaplastic, were used for validation. The classification performance of each approach is presented as the average over all trials in terms of Accuracy, Sensitivity and Specificity. The details for unsupervised and supervised feature learning methods are described next:

- For sAE, two different models were trained ( $sAE_{F:8,P:1}^{225}$  and  $sAE_{F:8,P:2}^{225}$ ) with 225 features and feature kernel size of  $8 \times 8$  only varying the pooling size by 1 and 2 respectively.
- For TICA, three different models were trained ( $TICA_{F:8,P:1}^{100}$ ,  $TICA_{F:16,P:1}^{100}$  and  $TICA_{F:8,P:1}^{225}$ ) using a pool size of 1 for all of them, while two used 100 features varying feature kernel size by  $8 \times 8$  and  $16 \times 16$  and the remaining model using 225 features with a feature kernel size of  $8 \times 8$ .
- For CNN, two models were trained ( $CNN_{F:8-8,P:2-2}^{CP16-CP32-FC128}$  and  $CNN_{F:8,P:2}^{CP225-FC225}$ ), the first is a 3-layers architecture with 16 features in the first layer, 32 features in the second layer, and 128 neurons in the full-connected layer, whereas the second model is a 2-layers architecture with 225 features in the first layer and 225

neurons in the full-connected layer. In both CNN, the feature kernel size was of  $8 \times 8$  and pool kernel of  $2 \times 2$ .

### 3.2.4 Results and discussion

Table **3-3** shows the results of medulloblastoma classification performance for each method in descending order of accuracy. It is clear that TICA, in all its configurations, achieve the best results. This suggests that topographic constraints improve the learned representation thanks to their invariant properties of scale and rotation in contrast to sAE. This has been also observed in previous works on another histopathology image analysis tasks different from tumor detection in basal cell carcinoma images [5].

The best results obtained for this tasks was achieved by  $TICA_{F:8,P:1}^{225}$  with 97% accuracy, 98% of sensitivity and 97% specificity, which outperforms by 10% the best previously reported results [42, 69].

Interestingly, CNN model, which had yielded very good results for natural scene image classification tasks, yields lower performance than TICA for  $CNN_{F:8,P:2}^{CP225-FC225}$  model, and even lower than sAE the best baseline results for  $CNN_{F:8-8,P:2-2}^{CP16-CP32-FC128}$ , in spite of the fact that it is a supervised learning method.

Finally, it is interesting to note that increasing the number of layers for CNN does not improve its performance. In fact, it looks like the most important factor for improving overall classification performance is to increase the number of features, even for shallow architectures. This suggests that pooling layers, which performed a spatial subsampling, are not as useful for the problem considered in this work since the object of interest (i.e. tumor type, anaplastic or non-anaplastic) tends to comprise the whole image, while in natural scene images the object of interest (e.g. car or person) appears only in a specific regions within the whole image. Hence, local spatial invariance provided by pooling layers is not particularly useful in our case.

### 3.2.5 Conclusions

In this paper we presented a comparative evaluation of supervised and unsupervised representation based learning methods for the problem of distinguishing anaplastic and non-anaplastic medulloblastoma from digitized histopathology. This comparative evaluation reveals some interesting findings of representation learning methods in the context of digital pathology related problems.

Deeper architectures do not necessarily produce better performance, experimental results showed a better performance for shallow architectures with more neurons. This suggests that the pooling layers for subsampling are not useful, contrary to the trend observed for natural scene images. Basically because the pooling layers provide local



Table **3-3**: Medulloblastoma classification performance comparison between unsupervised and supervised feature learning approaches.

Method	Accuracy	Sensitivity	Specificity
$TICA_{F:8,P:1}^{225}$	<b>0.97</b>	<b>0.98</b>	<b>0.97</b>
$TICA_{F:8,P:1}^{100}$	0.92	0.88	0.96
$TICA_{F:16,P:1}^{100}$	0.91	0.86	0.95
$CNN_{F:8,P:2}^{CP225-FC225}$	0.90	0.89	0.90
$sAE_{F:8,P:1}^{225}$	0.90	0.87	0.93
$sAE_{F:8,P:2}^{225}$	0.89	0.86	0.92
$CNN_{F:8-8,P:2-2}^{CP16-CP32-FC128}$	0.85	0.97	0.74
$BOF^{320} + A2NMF$ (Haar) [42]	0.87	0.86	0.87
$BOF^{320} + A2NMF$ (Block) [42]	0.78	0.89	0.67
$BOF + K - NN$ (Haar) [69]	0.80	-	-
$BOF + K - NN$ (MR8) [69]	0.62	-	-

space invariance, a property that is not sorely required for the problem considered. A reason may be that the object of interest (tumor type) does not appear in small regions of the image, in contrast with natural scene images. In this context, learning invariant feature representations based on topographic constraints produce better results since these features better capture scale and rotation invariance. This suggests that features that capture relevant information independent of scale and/or rotation are more discriminating.

In addition, this comparative evaluation corroborates what has been seen in other application areas, representation learning approaches obtain competitive results, and usually better, when they are compared against other data-driven representation methods such as bag of features. Future work will include a more exhaustive experimentation of deeper architectures for unsupervised and supervised approaches, evaluating interpretation capabilities of both approaches, evaluate methods that use context information (multi-resolution or multi-view), and evaluate them in other histopathology applications.

## 4 Supervised histopathology image representation learning

*Breast cancer is the most common type of cancer in women and the second cause of death in developed countries, whereas in Colombia it is the main cause of death by cancer. Precise invasive tumor delineation in the pathology slide is the very first step for a diagnosis workflow and allows further analysis of tumor differentiation via the well-known Bloom-Richardson and Nottingham grading schemes. These grading schemes estimate the grade of aggressiveness by evaluating histologic characteristics including: tubule formation, nuclear pleomorphism and mitotic count. This chapter presents a set of methods for mitosis detection (see into Section 4.1) and invasive tumor detection (see into Section 4.2) in breast cancer for digitized high-power fields and whole-slide images, based on supervised representation learning. The following sections correspond to papers presenting the details of the methods and applications.*

### 4.1 Mitosis detection in Breast Cancer

Breast cancer (BCa) grading plays an important role in predicting disease aggressiveness and patient outcome. A key component of BCa grade is mitotic count, which involves quantifying the number of cells in the process of dividing (i.e. undergoing mitosis) at a specific point in time. Currently mitosis counting is done manually by a pathologist looking at multiple high power fields on a glass slide under a microscope, an extremely laborious and time consuming process. The development of computerized systems for automated detection of mitotic nuclei, while highly desirable, is confounded by the highly variable shape and appearance of mitoses. Existing methods use either handcrafted features that capture certain morphological, statistical or textural attributes of mitoses or features learned with convolutional neural networks (CNN). While handcrafted features are inspired by the domain and the particular application, the data-driven CNN models tend to be domain agnostic and attempt to learn additional feature bases that cannot be represented through any of the handcrafted features. On the other hand, CNN is computationally more complex and needs a large number of labeled training instances. Since handcrafted features attempt to

model domain pertinent attributes and CNN approaches are largely supervised feature generation methods, there is an appeal to attempting to combine these two distinct classes of feature generation strategies to create an integrated set of attributes that can potentially outperform either class of feature extraction strategies individually. In this paper, we present a cascaded approach for mitosis detection that intelligently combines a CNN model and handcrafted features (morphology, color and texture features). By employing a light CNN model, the proposed approach is far less demanding computationally, and the cascaded strategy of combining handcrafted features and CNN-derived features enables the possibility of maximizing performance by leveraging the disconnected feature sets. Evaluation on the public ICPR12 mitosis dataset that has 226 mitoses annotated on 35 High Power Fields (HPF, 400x magnification) by several pathologists and 15 testing HPFs yielded an F-measure of 0.7345. Our approach is accurate, fast and requires fewer computing resources compared to existent methods, making this feasible for clinical use. *The complete content of this section has been published as a research article in the **Journal of Medical Imaging** (see [159]), the results of our participation in the **Assessment of Mitosis Detection Algorithms Challenge (AMIDA 2013)** were published in the **Journal of Medical Image Analysis** (see [156]) and preliminary work had been published in **Digital Pathology Conference. SPIE Medical Imaging 2014** (see [158]).*

#### 4.1.1 Introduction

Bloom Richardson grading [19], the most commonly used system for histopathologic diagnosis of invasive breast cancers (BCa) [62], comprises three main components: tubule formation, nuclear pleomorphism, and mitotic count. Mitotic count, which refers to the number of dividing cells (i.e. mitoses) visible in hematoxylin and eosin (H & E) stained histopathology, is widely acknowledged as a good predictor of tumor aggressiveness [71]. In clinical practice, pathologists define mitotic count as the number of mitotic nuclei identified visually in a fixed number of high power fields (400x magnification). However, the manual identification of mitotic nuclei often suffers from poor inter-interpreter agreement due to the highly variable texture and morphology between mitoses. Additionally this is a very laborious and time consuming process involving the pathologist manually looking at and counting mitoses from multiple high power view fields on a glass slide under a microscope. Computerized detection of mitotic nuclei will lead to increased accuracy and consistency while simultaneously reducing the time and cost needed for BCa diagnosis [73].

The detection of mitotic nuclei in H & E stained histopathology is a difficult task (see Fig. 4-1). First, mitosis is a complex biological process during which the cell nucleus

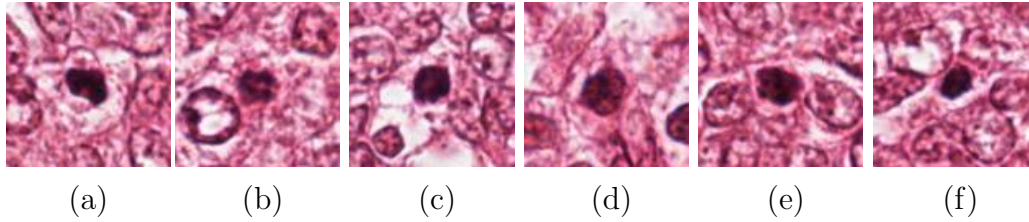


Figure 4-1: An illustration of the visual similarity between true mitotic processes and confounding false positives. (a)-(c): true mitoses. (d)-(f): confounding non-mitotic figures.

undergoes various morphological transformations. This leads to highly variable size and shape across mitotic nuclei within the same image. Another issue is rare event detection, which complicates classification tasks where one class (i.e. mitotic nuclei) is considerably less prevalent than the other class (i.e. non-mitotic nuclei).

Recently, the development of computerized systems for automated mitosis detection has become an active area of research with the goal of developing decision support systems to be able to relieve the workload of the pathologist. In a contest held in conjunction with the ICPR 2012 conference <sup>1</sup> [142] to identify the best automated mitosis detection algorithm, a variety of approaches competed against each other in the contest, which can be categorized as handcrafted feature based or feature learning based. The commonly used handcrafted features include various morphological, shape, statistical and textural features that attempt to model the appearance of the domain and in particular the appearance of the mitoses within the digitized images [82, 86, 87, 149].

While domain inspired approaches (hand crafted) are useful in that they allow for explicit modeling of the kinds of features that pathologists look for when identifying mitoses, there is another category of feature generation inspired by convolutional neural networks (CNN) [33, 102], CNN are multi-layer neural networks that learns a bank of convolutional filters at each layer [28, 116]. In contrast to handcrafted features, CNN is fully data-driven, therefore being more accurate in representing training samples and able to find feature patterns that handcrafted features fail to describe. However, CNN is computationally demanding and sensitive to the scalability of training data. The winner [28] of the ICPR contest used two 11-layers to achieve an F-measure of 0.78. However, this approach is not feasible for clinical use since each layer of the CNN model comprised hundreds of neurons and required a large amount of time (several weeks) for both training and testing.

<sup>1</sup><http://ipal.cnrs.fr/ICPR2012/>

Other methods achieved an F-measure of up to 0.71, based primarily on combining various handcrafted features. While handcrafted feature approaches are faster, drawbacks include (1) the fact that the identification of salient features are highly dependent on the evaluation dataset used and (2) the lack of a principled approach for combining disparate features. Hence, it stands to reason that a combination of CNN and handcrafted features will allow us to exploit the high accuracy of CNN while also reducing the computational burden (in terms of time) of training deep CNN models. By employing a light CNN model, the proposed approach is far less demanding computationally, and the cascaded strategy of combining handcrafted features and CNN-derived features enables the possibility of maximizing performance by leveraging the disconnected feature sets. Previous work in this approach includes the NEC team [116], where an attempt was made to stack the CNN-learned features and handcrafted features yielded an F-measure of 0.659, suggesting that more intelligent combination of CNN and handcraft features are required.

In this paper, we present a cascaded approach to combining CNN and handcrafted features for mitosis detection. The workflow of the new approach is depicted in Figure 4-2. The first step is to segment likely mitosis regions. This initial phase serves as a triage to remove obviously non-mitotic regions. For each candidate region, both CNN-learned and handcrafted features were extracted independently. Independently trained classifiers were constructed using the handcrafted and CNN-learned features alone. For the regions on which the two individual classifiers highly disagree, they are further classified by a third classifier that was trained based on the stacking of handcrafted and CNN-learned features. The final prediction score is a weighted average of the outputs of all the classifiers.

Our approach differs from the NEC system in two key aspects. First, we perform classification via CNN and handcrafted features separately, only using their combination to deal with confounders. Simply stacking handcrafted and CNN features will bias the classifier towards the feature set with the larger number of attributes. Our approach is less prone to this issue. Secondly, CNN works on a  $80 \times 80$  patch size while handcrafted features are extracted from clusters of segmented nuclei (normally  $\leq 30 \times 30$ ). This way we capture attributes of not only mitotic nuclei, but also its local context. Local context around candidate mitoses is an important factor for pathologists in correctly identifying mitoses. In summary, key novel contributions of this work include:

- A cascaded approach for combination of CNN and handcrafted features,
- Learning multiple attributes that characterize mitosis via the combination of CNN and handcrafted features,
- Achieving a high level of mitosis detection while minimizing the computing re-

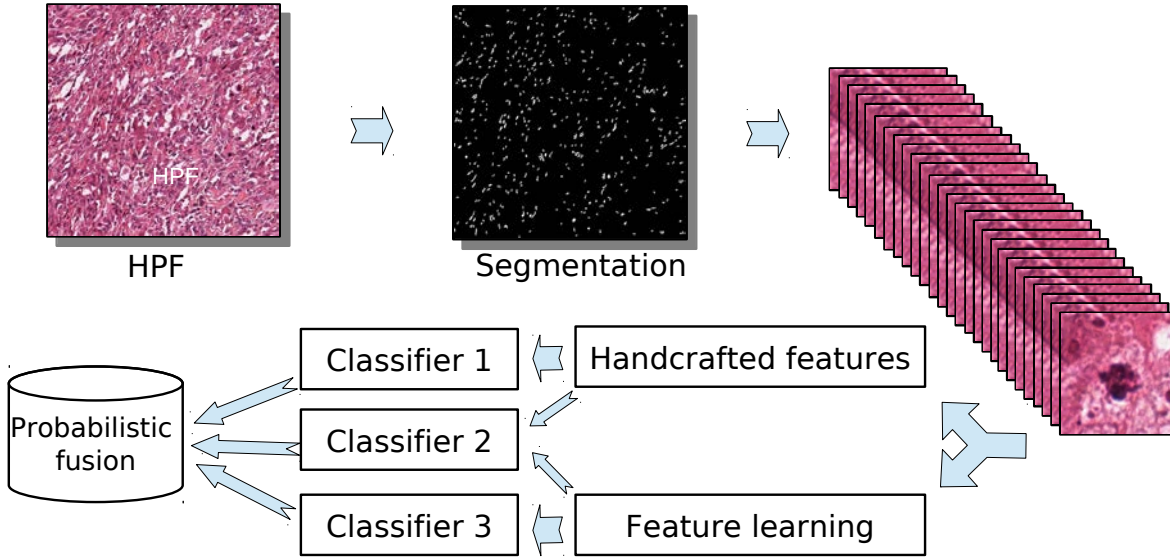


Figure 4-2: Workflow of our methodology. Blue-ratio thresholding [23] is first applied to segment mitosis candidates. On each segmented blob, handcrafted features are extracted and classified via a Random Forests classifier. Meanwhile, on each segmented  $80 \times 80$  patch, convolutional neural networks (CNN) [102] are trained with a fully connected regression model as part of the classification layer. For those candidates that are difficult to classify (ambiguous result from the CNN), we train a second-stage Random Forests classifier on the basis of combining CNN-derived and handcrafted features. Final decision is obtained via a consensus of the predictions of the three classifiers.

sources required.

## 4.1.2 Methodology

### Candidate Segmentation

We segment likely mitosis candidates by first converting RGB images into blue-ratio images [23]. By assigning a higher value to a pixel with a high blue intensity relative to its red and green components, blue-ratio is proven capable of highlighting nuclei regions [23]. Laplacian of Gaussian (LoG) [75] responses are then computed to discriminate the nuclei region from the background, followed by integrating globally fixed thresholding and local dynamic thresholding to identify candidate nuclei. One segmen-

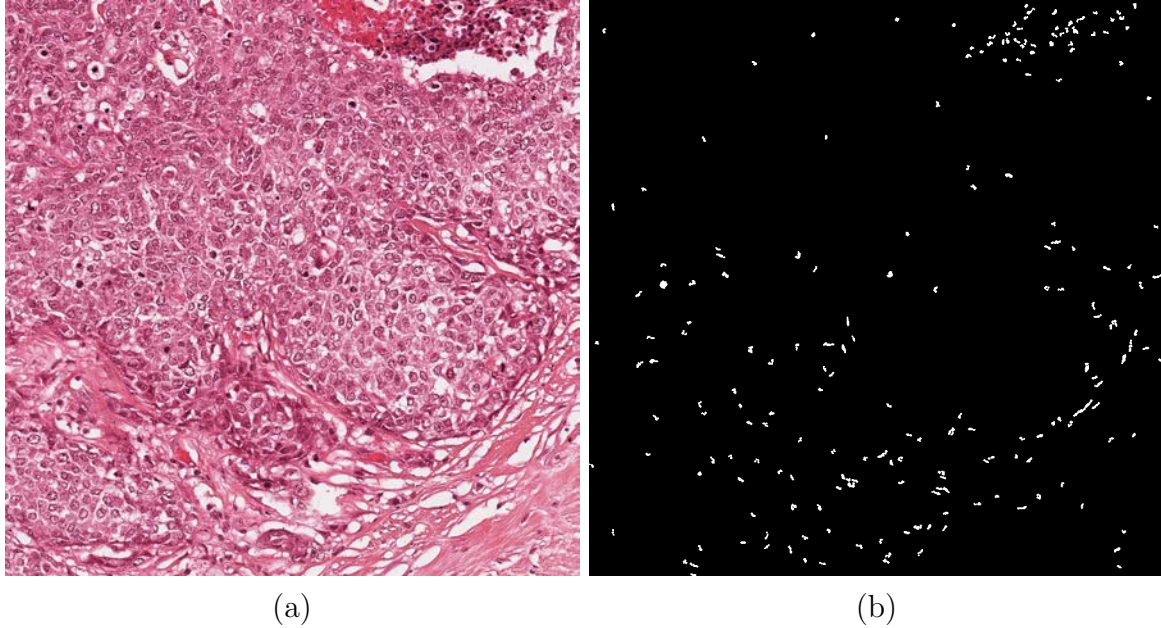


Figure 4-3: Example of blue-ratio segmentation. (a) is the original HPF slice while (b) is the segmentation mask. Note that a majority of the objects identified via this approach in (b) are indeed mitotic figures.

tation example is shown in Fig. 4-3. We can see that most dark-blue spots are retained as potential mitotic figures.

### Detection with Convolutional Neural Networks

#### CNN architecture

First, each HPF is converted from the RGB space to the YUV space and normalized to a mean of zero and variance of one. The CNN architecture employs 3 layers (Fig. 4-4): two consecutive convolutional and pooling layers and a final fully-connected layer. The convolution layer applies a 2D convolution of the input feature maps and a convolution kernel. The pooling layer applies a L2 pooling function over a spatial window without overlapping (pooling kernel) per each output feature map. Learning invariant features will be allowed through the L2 pooling. The output of the pooling layer is subsequently fed to a fully-connected layer, which produces a feature vector. The outputs of the fully-connected layer are two neurons (mitosis and non-mitosis) activated by a logistic regression model. The 3-layer CNN architecture comprises 64, 128, and 256 neurons, respectively. For each layer, a fixed  $8 \times 8$  convolutional kernel and  $2 \times 2$  pooling kernel were used.

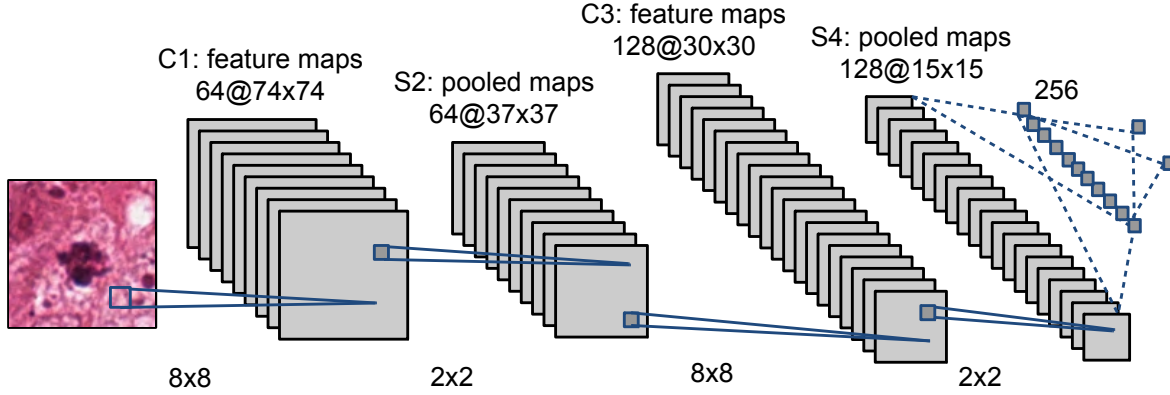


Figure 4-4: Architecture of the CNN model. The CNN architecture comprises 3 layers: two consecutive convolutional-pooling layers and a fully-connected classification layer. The two convolutional-pooling layers use the same fixed  $8 \times 8$  convolutional kernel and  $2 \times 2$  pooling kernel, but have 64 and 128 neurons, respectively. The last layer has 256 neurons, which are all connected to the final two neurons for mitosis/non-mitosis classification.

### Training stage

To deal with class-imbalance and achieve rotational invariance, candidate image patches containing mitotic nuclei were duplicated with artificial rotations and mirroring. The whole CNN model was trained using Stochastic Gradient Descent [20] to minimize the loss function:  $L(x) = -\log \left[ \frac{e^{x_i}}{\sum_j e^{x_j}} \right]$ , where  $x_i$  corresponds to outputs of a fully-connected layer multiplied by logistic model parameters. Thus the outputs of CNN are the log likelihoods of class membership.

### Testing stage

An exponential function is applied to the log likelihoods of each candidate nucleus belonging to the positive (mitosis) class in order to calculate the probability that it is mitotic. In our experiments, a candidate nucleus is classified as mitosis if the probability is larger than an empirically-determined threshold of 0.58.

## Detection with handcrafted features

### Features and Their Selection

The handcrafted features can be categorized into three groups: morphology, intensity and texture (Table 4-1). The morphological features are extracted from binary mask of mitosis candidate, which is generated by blue-ratio thresholding [23] and local



Table 4-1: Brief description of handcrafted features used for mitosis detection.

Category	Length	Features
Morphology	15	Area, eccentricity, equiv diameter, euler number, extent, perimeter, solidity, major axis length, minor axis length, area overlap ratio, average radial ratio, compactness, hausdorff dimension, smoothness and standard distance ratio.
Intensity	$8 \times 7$	Mean, median, variance, maximum/minimum ratio, range, interquartile range, kurtosis and skewness of patch intensities at 7 color channels.
Texture	$26 \times 7$	<u>Concurrence features</u> : mean and standard deviation of 13 Haralick [76] gray-level concurrence features grabbed at four orientations; <u>Run-Length features</u> [70]: mean and standard deviation of gray-level run-length matrices at four orientations;

non-maximum suppression. The morphological features represent various attributes of mitosis shape. Intensity and textural features are extracted from seven distinctive channels of squared candidate patches (Blue-ratio, Red, Blue, Green, L in LAB and V, L in LUV) according to [87]. The intensity features capture statistical attributes of mitosis intensity and the texture features capture textural attributes of mitosis region. The total length of handcrafted features is  $15 + 8 \times 7 + 26 \times 7 = 253$ . We then perform dimensionality reduction with principal component analysis [134]. The best features are retained in pearson1901pca by keeping 98% of the total component variations.

### Class Balancing and Classifier

We correct for the classification bias that occurs due to the relatively small number of mitotic nuclei compared to non-mitotic nuclei. To train a balanced classifier, we (1) reduce non-mitotic nuclei by replacing overlapping non-mitotic nuclei with their clustered center; (2) oversample mitotic cells by applying the Synthetic Minority Over-sampling Technique (SMOTE) [26], and (3) use an empirically-selected threshold 0.58. For classification, a Random Forest classifier with 50 trees is used. Using more trees tends to cause overfitting while using less trees leads to low classification accuracy.

### Cascaded Ensemble

The cascaded ensemble consists of two stages (shown in Fig. 4-5). First, we perform classification with CNN and handcrafted features individually. During training, we denote via  $L_d$  and  $L_h$  the classification labels associated with using CNN and handcrafted

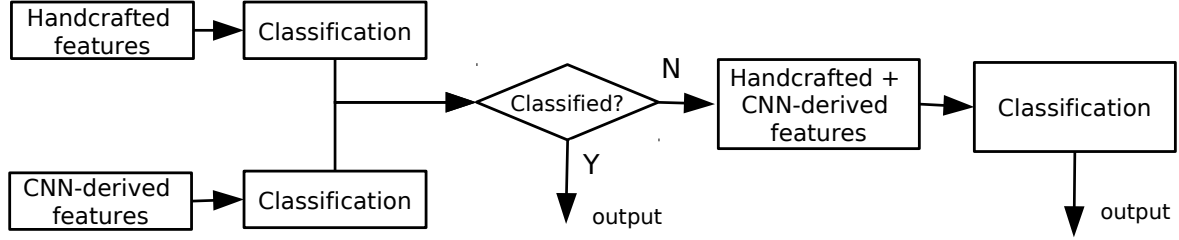


Figure 4-5: Workflow of the cascaded ensemble, which comprises two stages. First, we perform classification with CNN-learned and handcrafted features individually, and if the two classification scores are consistent, a binary decision (mitosis/non-mitosis) will be made directly. Secondly, for those instances whose individual classification scores are highly inconsistent, we classify them again by combining their CNN and handcrafted features.

features, respectively. For instances with  $L_d \neq L$  or  $L_h \neq L$ , where  $L$  is the ground truth label, we combine their CNN and handcrafted features to train a second-stage classifier  $\tilde{h}$ . During testing, given the output probabilities  $P_d$  and  $P_h$  of CNN and handcrafted feature classifiers, respectively, we calculate their combined probabilities  $P = w_d P_d + w_h P_h$ , where  $w_d$  and  $w_h$  are weighting factors. In the second stage, for instances with  $P \in [\lambda_l, \lambda_u]$  ( $\lambda_l$  and  $\lambda_u$  are certain lower and upper bounds, respectively), we let  $\tilde{h}$  classify them again. The instance having a final probability  $p$  larger than a certain threshold is categorized as mitosis, otherwise, non-mitosis.

### 4.1.3 Experimental design

#### ICPR Dataset

The dataset includes 50 images corresponding to 50 high-power fields (HPF) in 5 different biopsy slides stained with Hematoxylin and Eosin (H&E) (illustrated in Fig. 4-6). Each field represents a  $512 \times 512 \mu m^2$  area, and is acquired using three different setups: two slide scanners and a multispectral microscope. Here we consider images acquired by the widely-used Aperio XT scanner. The Aperio scanner has a resolution of  $0.2456 \mu m$  per pixel, resulting in a  $2084 \times 2084$  pixels RGB image for each field. A total of 326 mitotic nuclei are manually annotated by expert pathologist. The centroids of these mitoses are used as ground truth. According to the test, the first 35 HPF images (226 mitosis) are used for training, while the remaining 15 HPF images (100 mitosis) for evaluation.

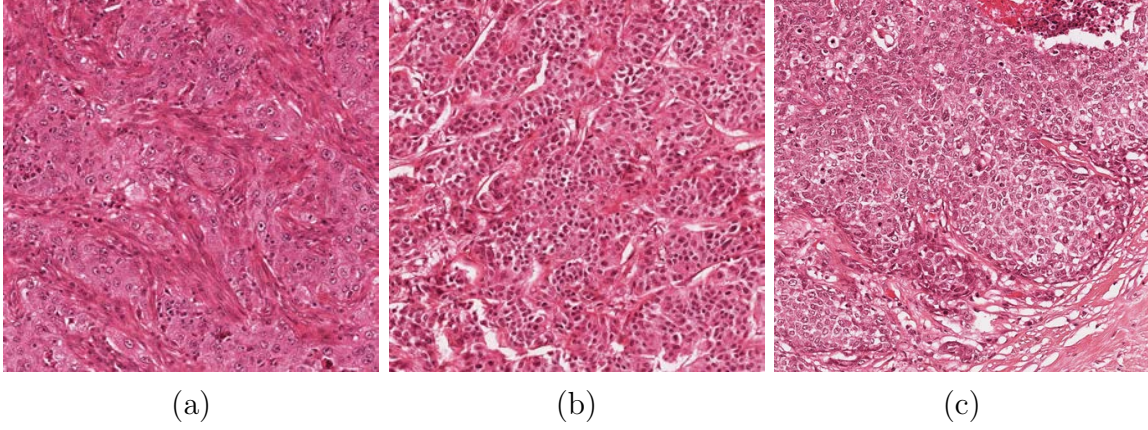


Figure 4-6: H&E-stained HPF examples from the ICPR dataset. The HPFs are acquired by a Aperio XT scanner with a resolution of  $0.2456\mu m$  per pixel. Each HPF has a size of  $2084 \times 2084$  pixels, representing a  $512 \times 512\mu m^2$  area annotated by pathologists.

### Performance Measures

Evaluation is performed according to the ICPR 2012 contest criteria, where true positives (TP) are defined as detected mitoses whose coordinates are closer than  $5\mu m$  (20.4 pixel) to the ground truth centroid. Nuclei that do not meet this criteria are defined as false positive (FP) and false negative (FN) errors. We compute the following performance measures:

$$Recall = \frac{TP}{TP + FN}, \quad Precision = \frac{TP}{TP + FP}, \quad F-measure = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (4-1)$$

We compare the proposed approach (**HC+CNN**) with approaches of using hand-crafted features only (**HC**), using CNN only (**CNN**), as well as the reported approaches in [142].

#### 4.1.4 Results and discussion

The mitosis detection results on ICPR12 dataset are shown in Table 4-2. The HC+CNN approach yields a higher F-measure (0.7345) than all other methods except that of ID-SIA (0.7821). The false negative rate associated with HC+CNN is relatively high as compared to other methods. As Table 4-3 illustrates, this is partially due to the fact that the blue-ratio segmentation has a false negative error of 7 mitoses. In addition,

Table 4-2: Evaluation results for mitosis detection using HC+CNN and comparative methods on the ICPR12 dataset.

Dataset	Method	TP	FP	FN	Precision	Recall	F-measure
Scanner Aperio	<b>HC+CNN</b>	65	12	35	0.84	0.65	<b>0.7345</b>
	<b>HC</b>	64	22	36	0.74	0.64	0.6864
	<b>CNN</b>	53	32	47	0.63	0.53	0.5730
	IDSIA [28]	70	9	30	0.89	0.70	0.7821
	IPAL [87]	74	32	26	0.70	0.74	0.7184
	SUTECH	72	31	28	0.70	0.72	0.7094
	NEC [116]	59	20	41	0.75	0.59	0.6592

Table 4-3: Performances of the blue-ratio segmentation module and the detection module. The blue-ratio segmentation finds 2484 mitosis candidates, among which 93 are true mitoses while the other 2391 are non-mitoses. 7 true mitoses are lost in this step. The detection module identifies 65 true mitoses and 12 false mitoses from these 2484 candidates. 28 mitoses are misclassified as non-mitotic figures in this module.

Segmentation module			Detection module			Final		
TP	FP	FN	TP	FP	FN	TP	FP	FN
93	2391	7	65	12	28	65	12	35

HC+CNN outperforms NEC (F-measure=0.6592), the only other approach to combine CNN and handcrafted features. Note that CNN based approaches (HC+CNN, IDSIA and NEC) tend to produce fewer FP errors, reflecting the capacity of CNN to accurately recognize non-mitotic nuclei.

The most critical parameter of the HC+CNN classifier is the classification threshold that is used to decide mitosis/non-mitosis. Based off our empirical results the optimal threshold was identified to be  $\approx 0.6$ . In general, a larger threshold will lead to less TPs, FPs and more FNs, and vice versa. In order to evaluate the influence of this threshold parameter, we generate the Precision-Recall (PR) curves by varying the threshold from 0.45 to 0.7. Fig. 4-7 shows that the performance of the other methods (except IDSIA) lie in the interior of the areas spanned by the curve. This fact suggests that the performance of HC+CNN is resilient to the precise choice of the classification threshold. Table 4-4 shows the influence of the number of random forests trees on mitosis detection. We can clearly see that fewer trees will most likely lead to worse

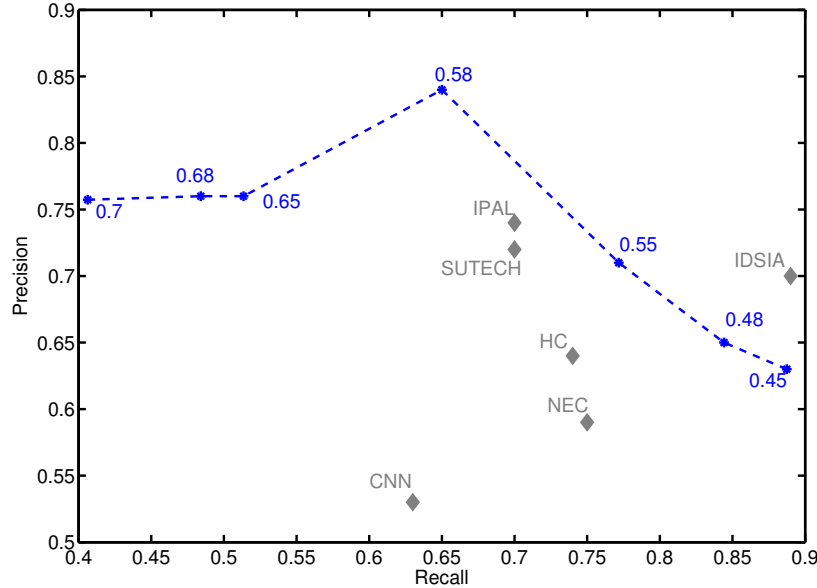


Figure 4-7: Precision-Recall (PR) curve of the proposed HC+CNN method. The performance of the other methods is also plotted for comparison. The curve is generated by varying the classification threshold between 0.45 and 0.7 (The threshold for each point is marked along the curve.). The fact that the performance of the other methods (except IDSIA) lie in the interior of the areas spanned by the curve suggests that the performance of HC+CNN is resilient to the precise choice of the classification threshold.

classification, while more trees may cause overfitting.

Figure 4-8 shows some detected mitosis examples. As one can see, the FNs tend to be poorly colored and textured while the FPs have similar color and shape attributes compared to the TPs. Although the textural patterns between FPs and TPs are different, this difference is not well appreciated at this pre-specified HPF resolution. Figure 4-9 shows a mitosis detection example using CNN and HC+CNN, respectively, revealing the improvement obtained by integrating handcrafted features and CNN in HC+CNN. Figure 4-10 show two mitotic detection results of HC+CNN, which also revealing some FN examples. Both the segmentation and detection steps contribute to the loss of these mitotic figures.

The two 11-layers neural networks used by IDSIA [28] requires roughly 30 epochs, which takes two days for training with GPU optimization. Our 3-layer CNN needs less than 10 epochs, and requires only 11.4 hours using 9 epochs without GPU op-

Table 4-4: The influence of the number of RF trees.

Number of Trees	10	20	30	50	100	200
F-measure	0.57	0.57	0.67	0.7345	0.65	0.66

timization. Including the time needed to extract handcrafted features (6.5 hours in pure MATLAB implementation), the training stage for HC+CNN was completed in less than 18 hours. At the detection stage, the Matlab implementation of HC+CNN takes about 1.5 minutes to process each H&E image, which is roughly 5x faster than the winner of the ICPR challenge [28].

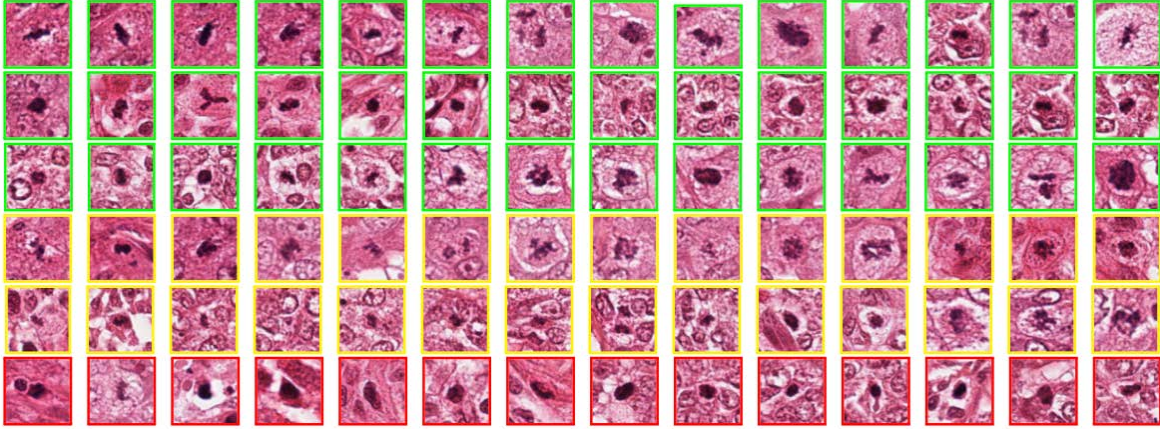


Figure 4-8: Mitoses identified by HC+CNN as TP (green rectangles), FN (yellow rectangles), and FP (red rectangles) on the ICPR12 dataset. The TP examples have distinctive intensity, shape and texture while the FN examples are less distinctive in intensity and shape. The FP examples are visually more alike to mitotic figures than the FNs.

#### 4.1.5 Conclusions

Mitosis detection is one of the three key factors in breast cancer grading. Existing approaches attempt to detect mitosis using either stacked handcrafted features or CNN-learned features. However, the problem of low detection accuracy arises when only handcrafted features are used while CNN-based approaches suffer from the issue of high computational complexity. To tackle these problems, we presented a new approach that combines handcrafted features and a light CNN in a cascaded way. Our approach yields a F-measure of 0.7345, which would have secured the second rank in the ICPR contest,



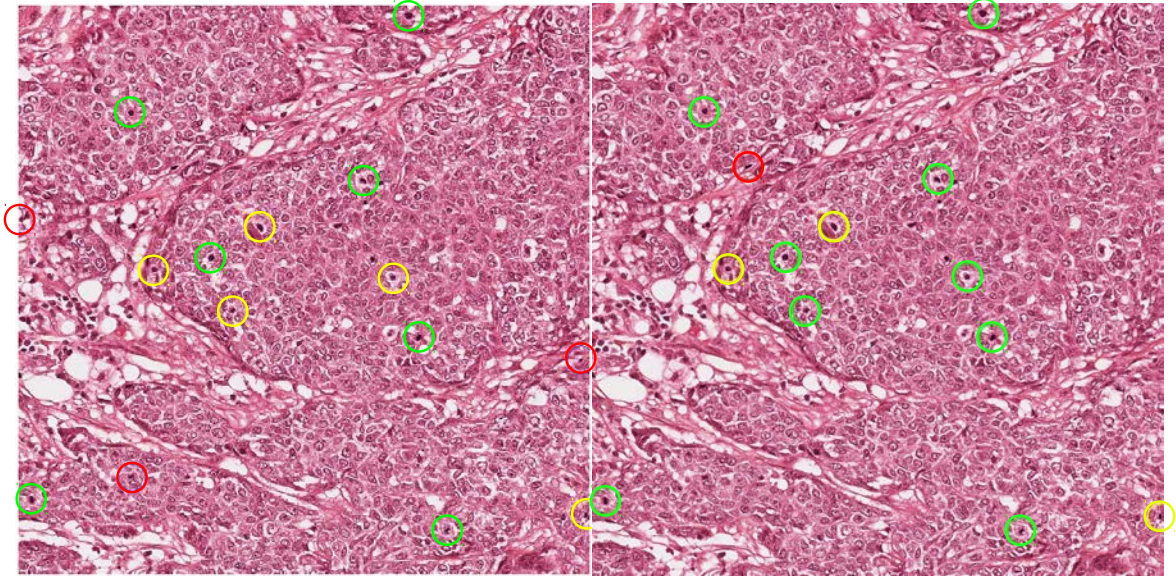


Figure 4-9: Mitoses identified by CNN and HC+CNN as TP (green circles), FN (yellow circles), and FP (red circles) on a HPF of ICPR12 dataset. On the left side, only using CNN leads to 7 TPs, 5 FNs and 3 FPs. On the right side, using HC and CNN leads to 9 TPs, 3 FNs and 1 FP, which clearly outperforms the use of CNN alone.

and higher than the NEC approach that combines CNN and handcrafted features at feature level. Compared to the leading methodology (two 11-layer CNN models) at the ICPR contest (F-measure = 0.78), our approach is faster, requiring far less computing resources.

Experimental results shows that it is still necessary to improve the accuracy of the presented approach. Future work will use GPU to implement a multi-layer (more than 3) CNN model.

## 4.2 Invasive BCa detection in Breast Cancer

With the increasing ability to routinely and rapidly digitize whole slide tissue images with slide scanners, there has been substantial interest in developing and applying computerized image analysis algorithms for automated detection and quantification of disease extent from digital pathology images. In the context of breast cancer, the manual identification of presence and extent of disease by a pathologist is critical for patient management in terms of tumor staging as well as assessing treatment response.

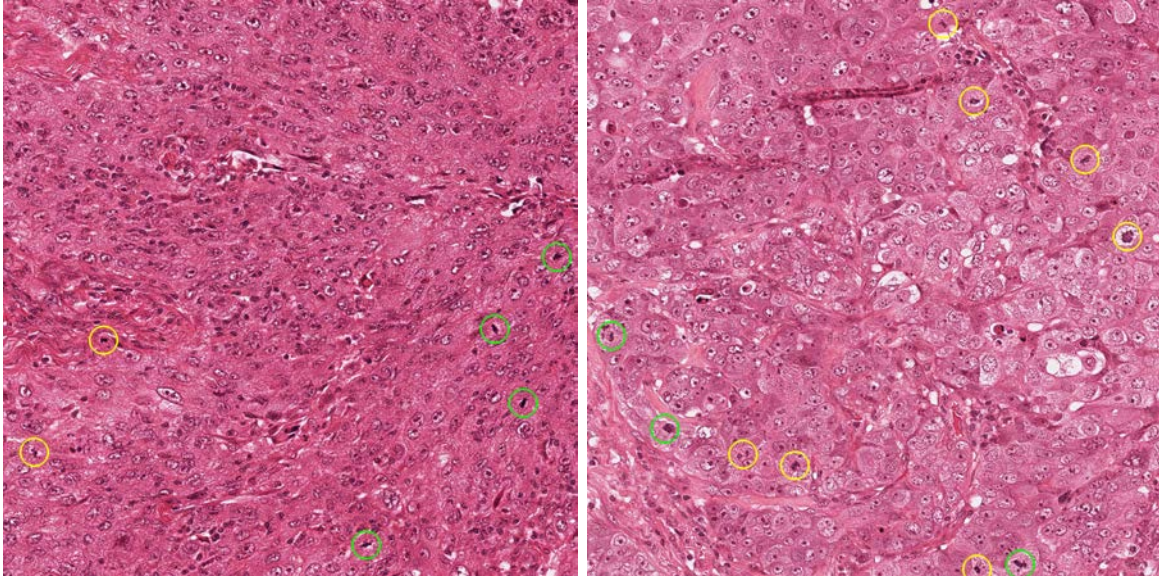


Figure 4-10: Mitoses identified by HC+CNN as TP (green circles) and FN (yellow circles) on two HPFs of the ICPR12 dataset. Mitoses on the left HPF have distinctive intensities and shapes, and confounding nuclei are few. Therefore, most mitoses can be correctly detected on this HPF. Comparatively, intensity of most mitotic nuclei on the right HPF is not distinctive enough for HC+CNN to identify, as a result, leading to a high FN.

However this process is often tedious and subject to inter- and intra-reader variability. Machine learning based image analysis methods offer the possibility of automatically detecting disease extent on digitized slide images. However, for these methods to be useful as decision support tools for the pathologist, they need to be resilient to data acquired from different sources, different staining and cutting protocols and different scanners. Therefore, the objective of this study was to quantitatively evaluate the accuracy and robustness of a deep learning based machine classifier to automatically identify the extent of invasive breast carcinoma on digitized whole slide images. In this study, we present a new machine learning method that employs a convolutional neural network for detecting presence and extent of invasive breast carcinoma on whole slide images. Our approach involves training the classifier on nearly 400 exemplars from multiple different sites, labs, and scanners, enabling the construction of an accurate and reproducible classifier, which was then independently validated on almost 200 cases from The Cancer Genome Atlas. Our approach yielded a Dice coefficient of 75.86%, a positive predictive value of 71.62% and a negative predictive value of 96.77% when compared against an expert pathologists annotation of cancer extent. Additionally, the



classifier yielded a negative predictive value of 100%, a true negative rate of 99.45% and a false positive rate of 0.55% in 11 normal control cases. Our results suggest that our approach could be employed as a decision support tool to aid pathologists in quantifying the extent and location of invasive breast cancer. . *The complete content of this section has been submitted as a research article in the **Journal of the American Medical Informatics Association** (see [41]), preliminary results had been published in **Digital Pathology Conference. SPIE Medical Imaging 2014** (see [36]) and an abstract was presented in **USCAP 2015** (see [35]).*

### 4.2.1 Introduction

Detection of tumor cells in a histologic section is the first step for the pathologist when diagnosing breast cancer (BCa). In particular, tumor delineation from background uninvolved tissue is a necessary prerequisite for subsequent tumor staging, grading and margin assessment by the pathologist [71]. However, precise tumor detection and delineation by experts is a tedious and time-consuming process, one associated with significant inter and intra-pathologist variability in diagnosis and interpretation of breast specimens [62,67,72,111,136]. Therefore, an automated and reproducible methodology for detection of invasive breast cancer on tissue slides could potentially reduce the total amount of time required to sign out a breast case and reduce some of this inter- and intra-observer variability [8,160]. Digital pathology refers to the process of digitization of tissue slides. The process of slide digitization could enable more efficient storage, visualization, and pathologic analysis of tissue slides and could potentially improve overall efficiency of routine diagnostic pathology workflow [113]. Quantitative histomorphometry refers to the application of computational image analysis and machine learning algorithms to identify and characterize disease patterns on digitized tissue slides [108]. In the context of breast cancer pathology, a number of computational imaging approaches have been recently applied for problems such as (i) detection of mitoses [28,87,112,156,159], tubules [11,44], nuclei [44,120], and lymphocytes [12], (ii) cancer grading [10,44], (iii) correlation of quantitative histologic image features and molecular features of breast cancer aggressiveness [9], and (iv) identification of histologic image features that are predictive of breast cancer outcome and survival [14]. These previous approaches have typically limited their analysis to only small portions of tissue or tissue microarrays (TMAs) as opposed to larger whole slide images. Basavanahally et al [10], looked at the problem of computerized Bloom-Richardson grading of estrogen receptor positive breast cancers within manually defined regions of interest on whole slide images. While some approaches have looked at the problem of classifying images as either containing cancer or not [60,124], no approach, that we are aware

of, has tackled the problem of automated delineation of invasive carcinoma on whole slide images. Neural network learning refers to a class of machine learning methods that is gaining popularity in histopathology image analysis [4, 5, 28, 31, 36, 100, 116, 159]. A neural network is composed of artificial neurons which are usually arranged in layers and interchange information through connections. In recent years, neural network models comprising thousands of neurons arranged in several layers have been shown to perform exceptionally well in computer vision and pattern analysis tasks [56, 77, 93, 99]. Multi-level neural network learning approaches have recently acquired the name ‘deep learning’ because of their multi-layer architecture. These networks are able to learn multiple levels of image representation in order to model complex relations among data, discovering more abstract and useful features that make it easier to extract useful information for high-level decision tasks such as segmentation, classification or prediction [15, 16, 47]. Because of the large number of parameters involved, deep learning methods require a large number of labeled training exemplars, in order to be optimally trained. In problems where large numbers of training exemplars are available, deep learning methods have shown impressive prediction results, often outperforming state-of-the-art classification methods [16, 93, 99]. The advent of digitized whole pathology slides and the concomitant increase in the number of publicly available large histopathology image databases, such as The Cancer Genome Atlas, has made digital pathology a good candidate for the application of deep learning based classification models [28, 33, 36, 100, 116, 159]. In this study, we present a classification approach for detecting presence and extent of invasive breast cancer on whole slide digitized pathology images using a ConvNet classifier [16, 102]. To ensure robustness of the classifier to variations in slide preparation, staining, and choice of scanning platform, we trained and validated the classifier with a large number of training exemplars drawn from three different institutions. Additionally the classifier was also independently evaluated on a large number of pathologic and normal cases drawn from The Cancer Genome Atlas (TCGA) and University Hospitals Case Medical Center. The goal of this study was to quantitatively evaluate the accuracy and robustness of a deep learning based machine classifier to automatically identify the extent of invasive breast cancer on digitized whole slide images.

## 4.2.2 Methodology

### Invasive Breast Cancer Tissue Detection in Whole-Slide Images

Our deep-learning based approach for detection of invasive breast cancer on whole-slide images is illustrated in Figure 4-11. The approach comprises three main steps: (i) tile tissue sampling, (ii) tile pre-processing, and (iii) convolutional neural network

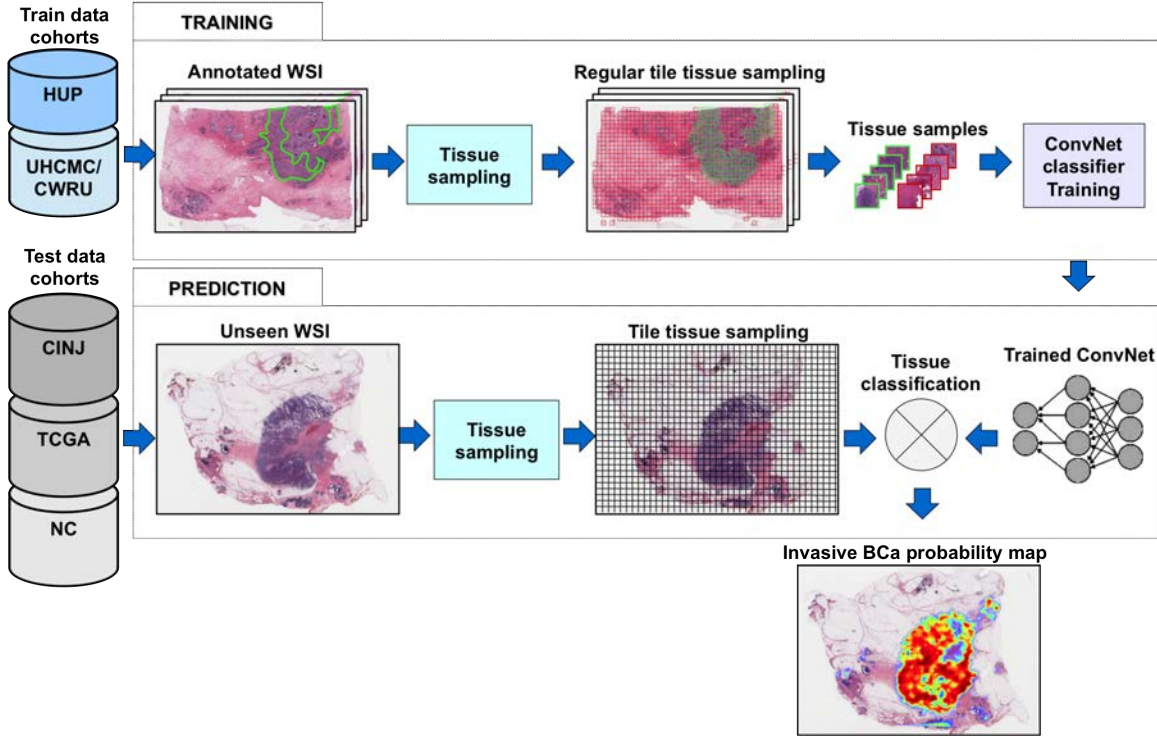


Figure 4-11: Overview of the process of training and testing of the deep learning classifiers for invasive breast cancer detection on whole-slide images.

(ConvNet) based classification. In this work, a tile is a square tissue region with a size of  $200 \times 200 \mu\text{m}$ . The tile tissue sampling process involves extraction of square regions of the same size ( $200 \times 200 \mu\text{m}$ ), on a rectangular grid for each whole-slide image. Only tissue regions are invoked during the sampling process and any regions corresponding to non-tissue within the background of the slide are ignored. The first part of the tile pre-processing procedure involves a color transformation from the original Red-Green-Blue color space representation to a YUV color space representation. A color normalization step is then applied to the digitized slide image to get zero mean and unit variance of the image intensities, and to remove correlations among the pixel intensity values. Tiles extracted from new whole-slide images, different from the ones used for training, are preprocessed using the same mean and standard deviation values in the YUV color space learnt during training. The ConvNet classifier [102], was trained using a set of image tiles extracted from invasive (positive examples) and non-invasive (negative examples) tissue regions, annotated on whole slide digitized images by expert pathologists. The ConvNet's architecture is comprised of 3-layers, the first layer is the

convolutional and pooling layer and the second is a fully connected layer, where each layer has 256 units (or neurons). The third is the classification layer with two units as outputs, one for each class (invasive and non-invasive), corresponding to a value between zero and one which can be interpreted as the probability of a given image tile being invasive cancer or not.

The ConvNet classifier was trained with images from HUP and UHCMC/CWRU. The best parameter configuration of the classifier was identified using the average area under the ROC curve (AUC) calculated over all slides in the CINJ data cohort,  $n=40$ . The AUC is a non-biased classification measure that allows for the evaluation of classification performance independent of a fixed threshold. In this work classification performance was evaluated over all the image tiles extracted from all the whole-slide images in the CINJ data cohort, tiles that correspond to either invasive or non-invasive tissue classes. This evaluation allows for the selection of a ConvNet classifier with the best classification performance (Avg. AUC =  $0.9018 \pm 0.0093$ ) for the subsequent experiments involving the independent test set.

### 4.2.3 Experimental design

#### Patients and Data Collection

This study involved images from five different cohorts from different institutions/pathology labs in the United States of America and TCGA [118,126]. The five cohorts were used for training, validation and independent testing of our method. The training data set had 349 estrogen receptor-positive (ER+) invasive breast cancer patients, of which 239 were from Hospital of the University of Pennsylvania (HUP), and 110 from University Hospitals Case Medical Center/Case Western Reserve University (UHCMC/CWRU). Patients from the HUP cohort ranged in age between 20 and 79 (average age  $55 \pm 10$ ). In the UHCMC/CWRU cohort, the patient age's ranged from 25 to 81 (average age  $58 \pm 10$ ). The validation data set contained 40 ER+ invasive breast cancer patients from the Cancer Institute of New Jersey (CINJ). The test data set was composed of two distinct subsets of positive and negative controls. For the test data set, we accrued a set of 195 ER+ invasive breast cancer cases from TCGA, age's ranging from 26 to 90 (average age  $57 \pm 13$ ). For the negative controls (NC) in the test data set, we used normal breast tissue sections taken from uninvolved adjacent tissue from 11 patients diagnosed with invasive ductal carcinoma from UHCMC/CWRU, Cleveland, OH. Patient specific information pertaining to race, tumor grade, and outcome were not explicitly recorded for this study.

Hematoxylin and eosin (H&E) slides were reviewed by the four study pathologists (NS, JT, MF, HG) to confirm the diagnosis of at least one type of invasive breast cancer

tumor. Tumors were categorized into one of the following histological types: Invasive carcinoma were categorized as either invasive ductal carcinoma (IDC) or invasive lobular carcinoma (ILC), while pre-invasive carcinoma was categorized as ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS).

### Slide Digitization and Pathologists Ground Truth

H&E stained histopathology slides were digitized via a whole-slide scanner at 40x magnification for this study. An Aperio Scanscope CS scanner was used to digitize cases from the HUP, CINJ and TCGA cohorts. The Ventana iCoreo scanner was used for scanning the UHCCMC/CWRU and NC data cohorts. 40x magnification corresponds to Aperio’s slides at  $0.25\ \mu\text{m}/\text{pixel}$  resolution and to Ventana’s slides at  $0.23\ \mu\text{m}/\text{pixel}$ . Expert pathologists provided the ground truth annotations of invasive breast cancer regions over digitized histopathology slides for all the data cohorts (HUP, CINJ, UHCCMC/CWRU, TCGA). The region annotations were obtained via manual delineation of invasive breast cancer regions by expert pathologists using the viewing software applications ImageScope v11.2 from Aperio and Ventana Image Viewer v3.1.4 from Ventana. To alleviate the time and effort required to create the ground truth annotations for extent of invasive breast cancer, the pathologists were asked to perform their annotations at 2x magnification or less. All whole-slide images previously sampled at 40x were thus subsequently downsampled (by a factor of 16:1) to a resolution of  $4\ \mu\text{m}/\text{pixel}$ .

### Method evaluation

We evaluated the accuracy of the ConvNet classifier in whole slide images by comparing the predictions of invasive regions in the test dataset against the corresponding ground-truth regions annotated by expert pathologists. The test data sets included the slides in the TCGA and NC cohorts. A quantitative evaluation was performed by measuring the Dice coefficient, positive predictive value (PPV), negative predictive value (NPV), true positive rate (TPR), true negative rate (TNR), false positive rate (FPR) and false negative rate (FNR) across all the test slides. These measures were evaluated for each whole-slide image and the mean and standard deviation in performance measures were calculated for each test data cohort.

In addition to training the ConvNet classifier with the full training data set (HUP and UHCCMC/CWRU), two additional classifiers were trained using, in each case, one of the training cohorts:  $\text{ConvNet}_{\text{HUP}}$  trained with the HUP cohort and  $\text{ConvNet}_{\text{UHCCMC/CWRU}}$  trained with the UHCCMC/CWRU cohort. The motivation was to analyze the sensitivity of the classifier to the training data sets. Both  $\text{ConvNet}_{\text{HUP}}$  and  $\text{ConvNet}_{\text{UHCCMC/CWRU}}$

were evaluated on both the validation (CINJ cohort) and test datasets (TCGA and NC cohorts) to analyze how and where their predictions diverged. Specifically we measured the correlation coefficient  $r$  between the prediction performance measures for  $\text{ConvNet}_{\text{HUP}}$  and  $\text{ConvNet}_{\text{UHCMC/CWRU}}$  across all slides in each test cohort.

#### 4.2.4 Results and discussion

##### Quantitative evaluation for automatic invasive breast cancer detection

Table 4-5 shows the detection performance in terms of mean and standard deviation of Dice coefficient, PPV, NPV, TPR, TNR, FPR and FNR through the validation dataset, comprised of the TCGA and the NC cohorts. Figure 4-12 shows some representative slide images from the validation dataset. Figures 4-12A-C depict the ground truth annotations from the pathologists on three whole-slide images from the TCGA data cohort and Figures 4-12D-F represent the automatic predictions of the fully-trained ConvNet classifier as a probability map of invasive breast cancer, with the color bar reflecting the probability values, high probability values reflected in red colors and low probability values in blue colors. Finally, three example slides without any malignant pathology and part of the NC cases are illustrated in Figures 4-12G-I. As may be seen in Figures 4-12G-I, the ConvNet classifier did not identify any regions as having invasive breast cancer.

Table 4-5: Performance measures for the ConvNet classifier on the TCGA (pathological) and NC (normal) data cohorts. The measures included Dice, PPV, NPV, TPR, TNR, FPR and FNR. Note that for the normal cases considered, not all the performance measures are shown because the NC data cohort did not have cancer annotations.

Dataset	Dice	PPV	NPV	TPR	TNR	FPR	FNR
TCGA	0.7586	0.7162	0.9677	0.8691	0.9218	0.0782	0.1309
	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	0.2006	0.2204	0.0511	0.1582	0.0764	0.0764	0.1582
NC	N/A	N/A	1 +/- 0	N/A	0.9945	0.0055	N/A
					+/-	+/-	
					0.0147	0.0147	

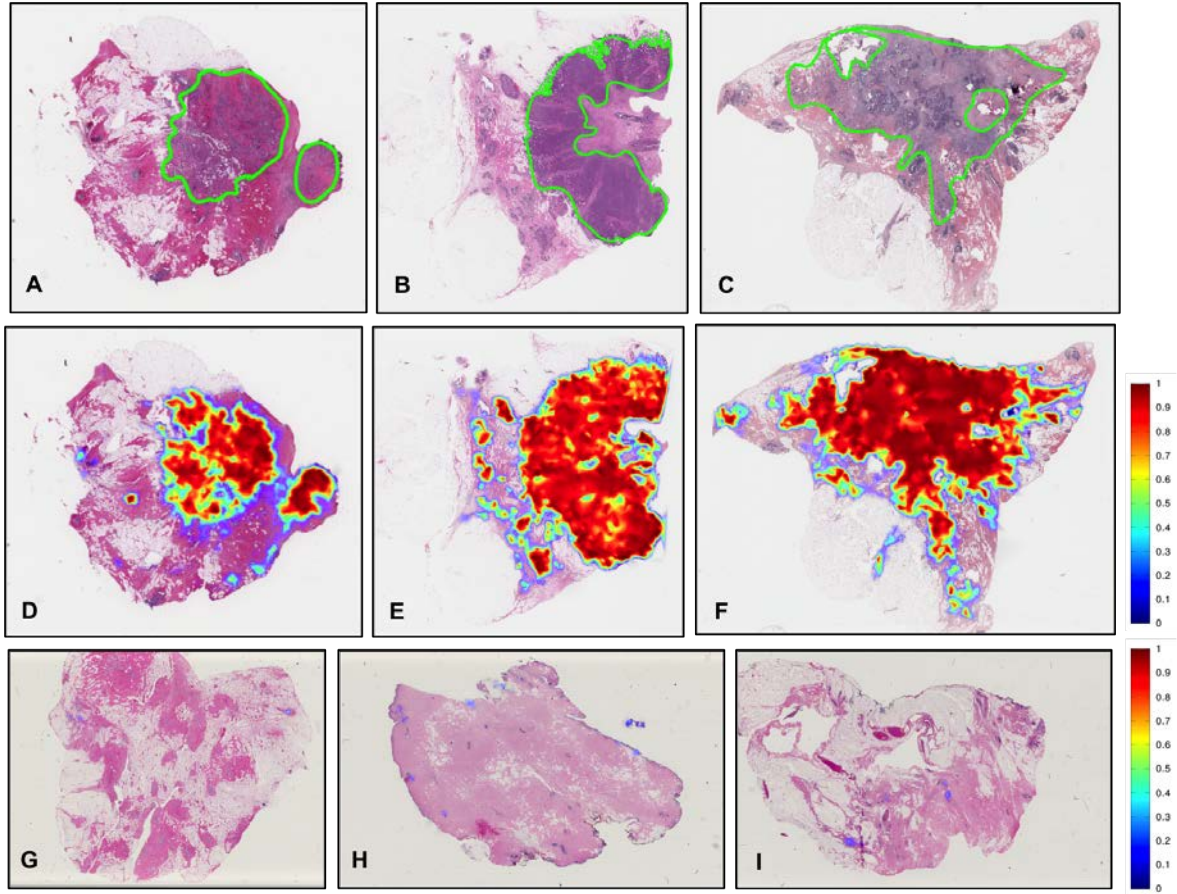


Figure 4-12: (A-C) Example whole-slide images from test TCGA data cohort with ground truth annotations from pathologists, (D-F) the corresponding region predictions produced by the ConvNet classifier and (G-I) region predictions for whole-slide images from the test NC data cohort of normal breast tissue without cancer.

### Robustness and reproducibility analysis inside heterogeneous histopathology slides

A detailed analysis by subgroups of only a type of invasive breast cancer (i.e. IDC or ILC) and mixture of invasive and other types of in situ lesions (e.g. DCIS and LCIS) is presented in Table 4-6 for each of ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> classifiers. Each of ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> was trained with one of either the HUP or the UHCMC/CWRU cohorts. The quantitative performance results for both classifiers, ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub>, on the validation CINJ data cohort

(ConvNet<sub>HUP</sub>: Dice = 0.6771, PPV = 0.6464, NPV = 0.9709; ConvNet<sub>UHCMC/CWRU</sub>: Dice = 0.6596, PPV = 0.6370, NPV = 0.9663) are similar. The results in Table 4-6 are also arranged according to the type of tumors in the sample (mixture or only invasive) and reveal that our method has better performance when the whole-slide images have only one type of invasive tumor (ConvNet<sub>HUP</sub>: Dice = 0.7578, PPV = 0.7462, NPV = 0.9654; ConvNet<sub>UHCMC/CWRU</sub>: Dice = 0.7596, PPV = 0.7462, NPV = 0.9614).

Table 4-6: Performance of the ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> classifiers on the CINJ data cohort in terms of means and standard deviation of Dice coefficient, PPV and NPV. The results in Table 4-6 are organized in terms of all cases in the CINJ cohort (N=40), a subset of the CINJ cohort with invasive breast cancer alone (N=19), and a mixture of invasive and other in situ subtypes of breast cancer (N=21).

ConvNet <sub>HUP</sub>				
Group	N	Dice	PPV	NPV
All cases	40	0.6771 +/- 0.2445	0.6464 +/- 0.2870	0.9709 +/- 0.0350
Only invasive	19	0.7578 +/- 0.2166	0.7462 +/- 0.2480	0.9654 +/- 0.0355
Mixture	21	0.6041 +/- 0.2501	0.5560 +/- 0.2953	0.5560 +/- 0.2953
ConvNet <sub>UHCMC/CWRU</sub>				
Group	N	Dice	PPV	NPV
All cases	40	0.6596 +/- 0.2527	0.6370 +/- 0.2941	0.9663 +/- 0.0421
Only invasive	19	0.7596 +/- 0.2074	0.7499 +/- 0.2423	0.9614 +/- 0.0440
Mixture	21	0.5691 +/- 0.2602	0.5348 +/- 0.3045	0.9708 +/- 0.0409

Figure 4-13 illustrates representative examples of whole slide images from the validation CINJ data cohort, involving only a single type of invasive tumor. The detection results obtained via ConvNet<sub>HUP</sub> classifier were compared against the ground truth annotations. Some cases from the CINJ validation data cohort where the ConvNet<sub>HUP</sub> classifier resulted in a poor detection performance are illustrated in Figures 4-14 and 4-15. The true-positives (TP), true-negatives (TN), false-positives (FP) and false-negatives (FN) regions, based on the predictions of the ConvNet<sub>HUP</sub> classifier, are illustrated in green, blue, yellow and red respectively. Figure 4-14 shows a case of mucinous (colloid) carcinoma which is a rare type of invasive ductal carcinoma with a very low prevalence (2-3% of the total invasive breast cancer cases) [96]. Figure 4-15 depicts a challenging case, which is composed of a mixture of invasive and in situ carcinoma elements.



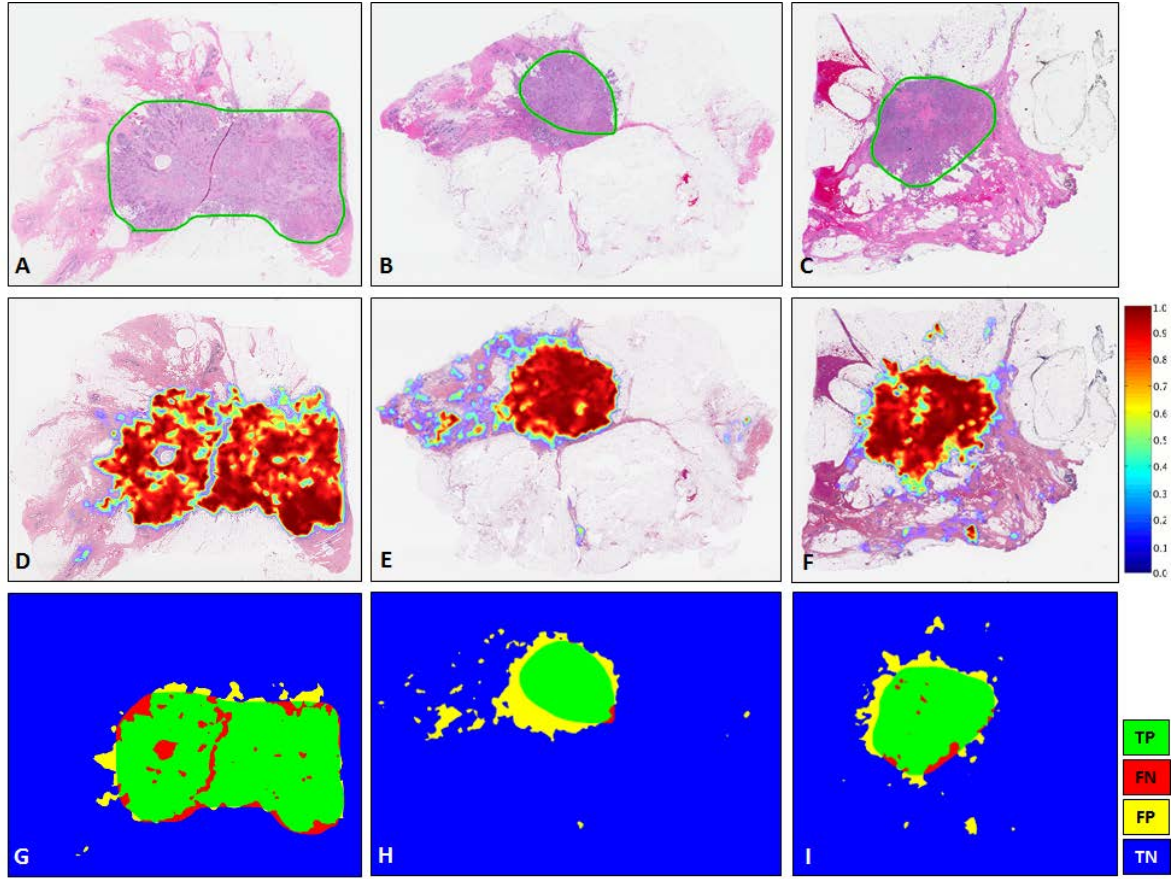


Figure 4-13: Example results for the ConvNet<sub>HUP</sub> classifier on the CINJ validation data cohort. The probability map predicted by the ConvNet<sub>HUP</sub> classifier (second row, D-F) was compared against ground truth annotations by a pathologist (first row, A-C). The third row shows the evaluation results of the ConvNet<sub>HUP</sub> classifier in terms of TP (green), FN (red), FP (yellow), and TN (blue) regions.

### Correspondence and reproducibility analysis among different classifiers and data cohorts

Table 4-7 presents the performance measures for the ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> classifiers on the TCGA and NC testing cohorts. The consistency of the predictions of both models is estimated by calculating the correlation coefficient,  $r$ , between the performance measures obtained for each of ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub>. On the TCGA cohort, the correlation coefficient for ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> in Dice coefficient was  $r=0.8733$ , reflecting a high degree of consistency between the

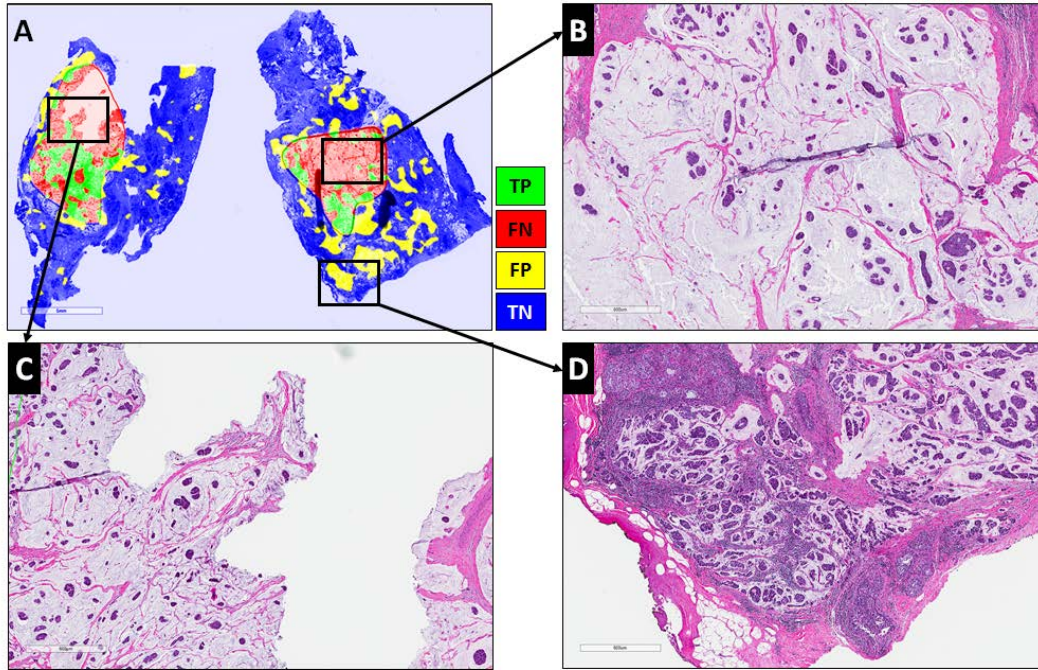


Figure 4-14: Whole-slide image from CINJ validation data cohort diagnosed with a rare type of IDC: mucinous carcinoma of the breast. (A) The comparison between the ground truth annotations and the predictions from the ConvNet<sub>HUP</sub> classifier reveal both FN (red) and FP (yellow) errors. (B-C) Most of the FN regions, i.e. tissues wrongly labeled as non-invasive tumor, correspond to mucinous carcinoma, whilst (D) most of FP regions, i.e. tissues wrongly predicted as invasive tumor, are actually invasive mucinous carcinoma that was not included in the annotations by the pathologist.

two classifiers trained on two different data cohorts. Figure 4-16 shows a scatter plot where the X axis corresponds to the Dice coefficient of the predictions generated by the ConvNet<sub>HUP</sub> and the Y axis corresponds to the Dice coefficient of the predictions generated by the ConvNet<sub>UHCMC/WRU</sub>, each dot correspond to a slide sample from the TCGA data cohort. The scatter plot reveals a well defined cluster with most of cases aggregating in the upper-right corner, allowing us to visually conclude that both ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> have a high degree of precision and agreement in their predictions. Figure 4-16 also allows us to identify some cases (red circles) where both ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> disagreed in their predictions. Figure 4-17 which showcases these images suggests that the discrepancy might be due to variations in slide staining. Figures 4-17A-B illustrate a couple of slides characterized



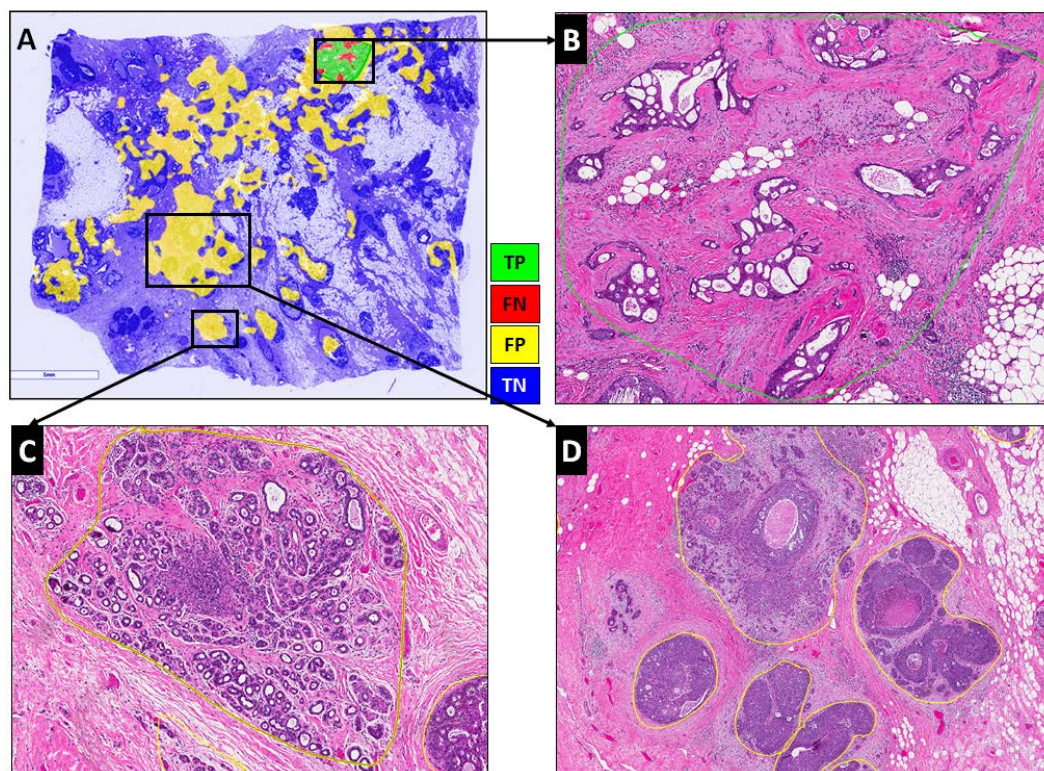


Figure 4-15: The most challenging whole-slide image in the CINJ validation cohort achieved the poorest performance via the  $\text{ConvNet}_{\text{HUP}}$  classifier with (A) many FP regions and a Dice coefficient of 0.0745. (B) Some of the FN errors are due to the confounding morphologic attributes of the tumor, arising due to a mixing of IDC with fat cells and irregular, infiltrating looking cribriform glands with DCIS. The FP regions appear to be primarily be due to (C) sclerosing adenosis, and (D) DCIS surrounded by IDC.

by low levels of hematoxylin and high levels of eosin. The slide shown in Figure 4-17C represents an example of a "black discoloration artifact" due to air bubbles on the slide, a common problem when the slide has been in storage for a long time. Usually, these cases are not appropriate for diagnosis and a pathologist would probably reject them in a quality control process requiring a repeat of the histopathology slide mounting process.

Despite these special cases of disagreement caused by staining issues, both the  $\text{ConvNet}_{\text{HUP}}$  and  $\text{ConvNet}_{\text{UHCMC/CWRU}}$  classifiers yielded similar predictions and performance. However, the  $\text{ConvNet}_{\text{HUP}}$  classifier appears to have a slightly higher confidence interval associated with the Dice and PPV performance measures. On the other hand, NPV

Table 4-7: Comparison and correlation of the ConvNet<sub>UHCMC/CWRU</sub> and ConvNet<sub>HUP</sub> classifiers in terms of Dice, PPV, NPV, TPR, TNR, FPR and FNR. Note that for the normal cases considered, not all the performance measures are shown because the NC data cohort did not have cancer annotations.

	<b>TCGA</b>						
	<b>Dice</b>	<b>PPV</b>	<b>NPV</b>	<b>TPR</b>	<b>TNR</b>	<b>FPR</b>	<b>FNR</b>
ConvNet <sub>HUP</sub>	0.7494	0.7071	0.9658	0.8600	0.9188	0.0812	0.1400
	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	0.2071	0.2254	0.0514	0.1705	0.0805	0.0805	0.1705
ConvNet <sub>UHCMC/CWRU</sub>	0.7068	0.6464	0.9629	0.8676	0.8880	0.1120	0.1324
	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	0.2061	0.2188	0.0584	0.1706	0.0824	0.0824	0.1706
$r$	0.8733	0.9258	0.8109	0.6345	0.8055	0.8055	0.6345
	<b>NC</b>						
	<b>Dice</b>	<b>PPV</b>	<b>NPV</b>	<b>TPR</b>	<b>TNR</b>	<b>FPR</b>	<b>FNR</b>
ConvNet <sub>HUP</sub>	N/A	N/A	1 +/- 0	N/A	0.9716	0.0284	N/A
					+/-	+/-	
					0.0693	0.0693	
ConvNet <sub>UHCMC/CWRU</sub>	N/A	N/A	1 +/- 0	N/A	0.9546	0.0454	N/A
					+/-	+/-	
					0.0816	0.0816	
$r$	N/A	N/A	N/A	N/A	0.6876	0.6876	N/A

and TNR from both classifiers show high mean values with very small standard deviation. Similarly on the NC data cohort, which is exclusively composed of normal breast samples, both the ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> classifiers exhibited a very high mean TNR and a very low FPR, with very low associated standard deviation. This appears to suggest that both classifiers are able to confidently and consistently reject non-invasive tissue regions.

Example results of the predictions from the ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> classifiers on the TCGA and NC test datasets are presented in Figures 4-18 and 4-19. While both the ConvNet<sub>HUP</sub> and ConvNet<sub>UHCMC/CWRU</sub> classifiers tend to produce consistent predictions, the ConvNet classifier which was trained using the complete training dataset, had the best overall performance (Figure 4-12).

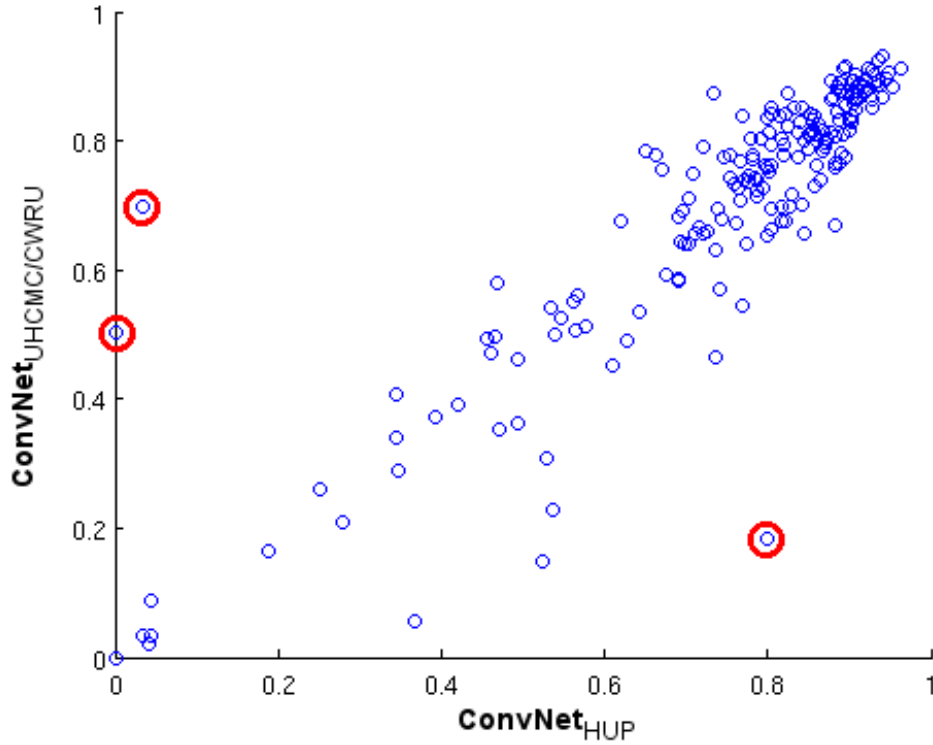


Figure 4-16: Agreement plot of the Dice coefficient for the  $\text{ConvNet}_{\text{HUP}}$  (X-axis) and  $\text{ConvNet}_{\text{UHCMC/CWRU}}$  (Y-axis) classifiers for each slide (blue circles) in the TCGA cohort. The slides with higher disagreement are identified with red circles (see Figure 4-17).

## Discussion

The experimental results show that the method is able to detect invasive breast cancer regions on whole slide histopathology images with a high degree of precision, even when tested on cases from a cohort different to the one used for training. The most challenging cases for the method were slides where invasive breast cancer was mixed in with in situ disease (which is not surprising and could be reduced by training a more complex network that included examples of these precursor lesions).

An important part of the experimental setup was the analysis of the detection sensitivity of the method to the data used for training. The results show that the classifiers trained with two different data cohorts, HUP and UHCMC/CWRU, exhibit highly correlated performance measures ( $r \geq 0.8$ ) over the independent TCGA test data cohort (see Table 4-7). Despite this, there are some differences in the prediction performance of the two classifiers, possibly suggesting "batch effects" [90], that originated

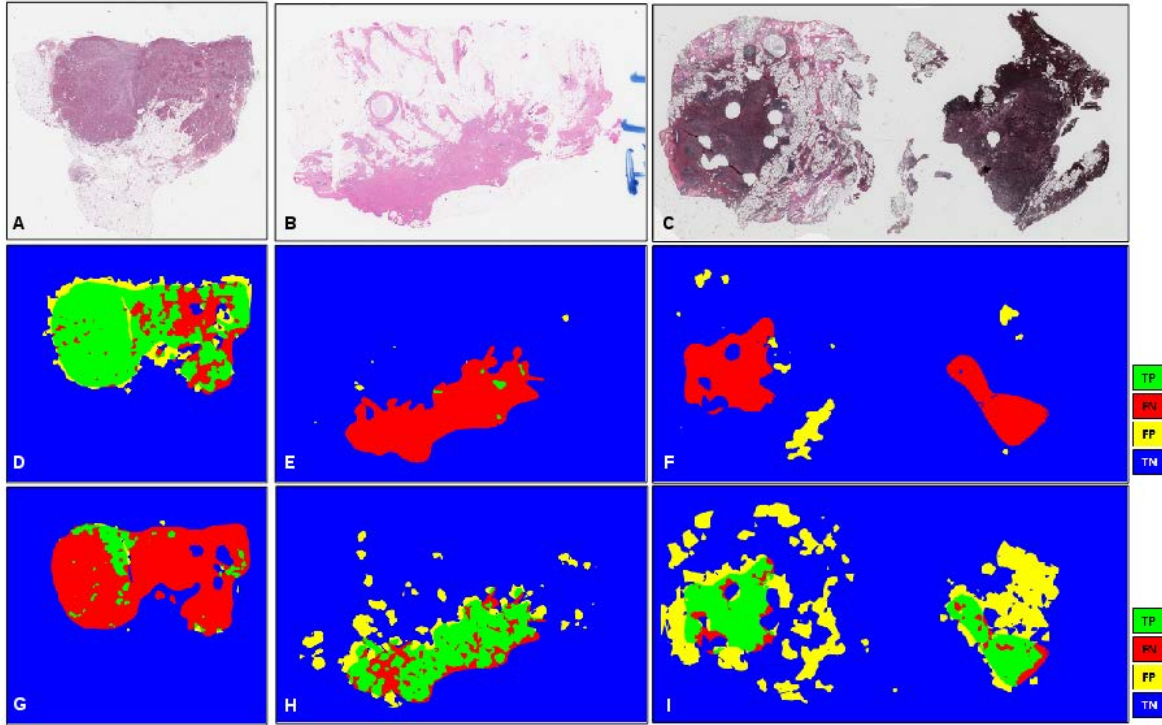


Figure 4-17: (A-C) Slides from from the TCGA cohort which revealed disagreement between the predictions of the  $\text{ConvNet}_{\text{HUP}}$  and  $\text{ConvNet}_{\text{UHCMC/CWRU}}$  classifiers. The predictions of the (D-F)  $\text{ConvNet}_{\text{HUP}}$  and (G-I)  $\text{ConvNet}_{\text{UHCMC/CWRU}}$  classifiers were compared against the ground truth annotations in terms of TP (green), FN (red), FP (yellow) and TN (blue) regions.

from the process of ground truth annotation or slide digitization. This is illustrated in Figures 4-16 and 4-17, which shows representative slides with artifacts due to problems in the histotechnique process. The method shows a very low false positive rate, this is evidenced by the results in the NC cohort ( $\text{ConvNet}_{\text{HUP}}$ :  $\text{FPR} = 0.0284$ ;  $\text{ConvNet}_{\text{UHCMC/CWRU}}$ :  $\text{FPR} = 0.0454$ ), which comprised only normal breast sections. The performance of machine learning methods improved as the number of training samples increased, with the  $\text{ConvNet}$  classifier which was trained with both the HUP and UHCMC/CWRU data cohorts yielding the best overall performance (Table 4-5 and Figure 4-12).

Our study did however have its limitations. There are some subtypes of invasive breast cancers that our method is not able to detect in a precise way such as the rare special histologic subtype mucinous carcinoma that comprises around 3% of the invasive breast



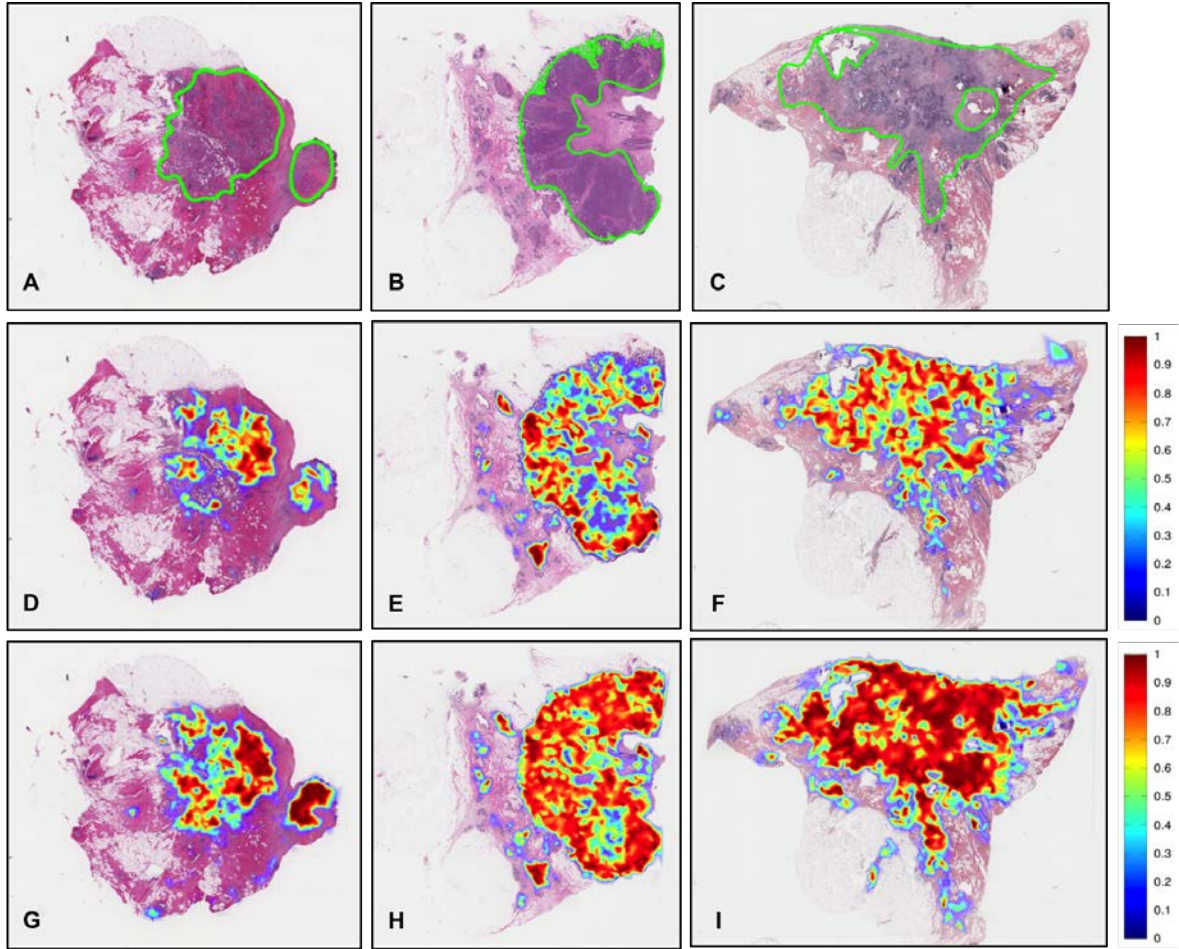


Figure 4-18: (A-C) Example whole-slide images from the TCGA data cohort with corresponding ground truth annotations. The probability maps generated by the  $\text{ConvNet}_{\text{UHMC/CWRU}}$  and  $\text{ConvNet}_{\text{HUP}}$  classifiers are shown in panels (D-F) and (G-I) respectively.

cancers. In fact, in the test dataset there are two cases similar to Figure 4-14, with mucinous carcinoma that were not detected. Another limitation is that some in situ breast cancer regions were incorrectly classified as invasive breast cancer, in situ disease is different from invasive cancer. However, the reporting of the presence of both invasive and in situ carcinoma is a critical part of a diagnostic pathology workup. It is worth noting though that our approach was able to achieve a very high level of accuracy in terms of rejecting non-invasive tissue regions (normal controls) as not being cancer. This is a remarkable fact considering that invasive tissue samples were the key training variable and that exemplars of DCIS and LCIS were not explicitly included in the

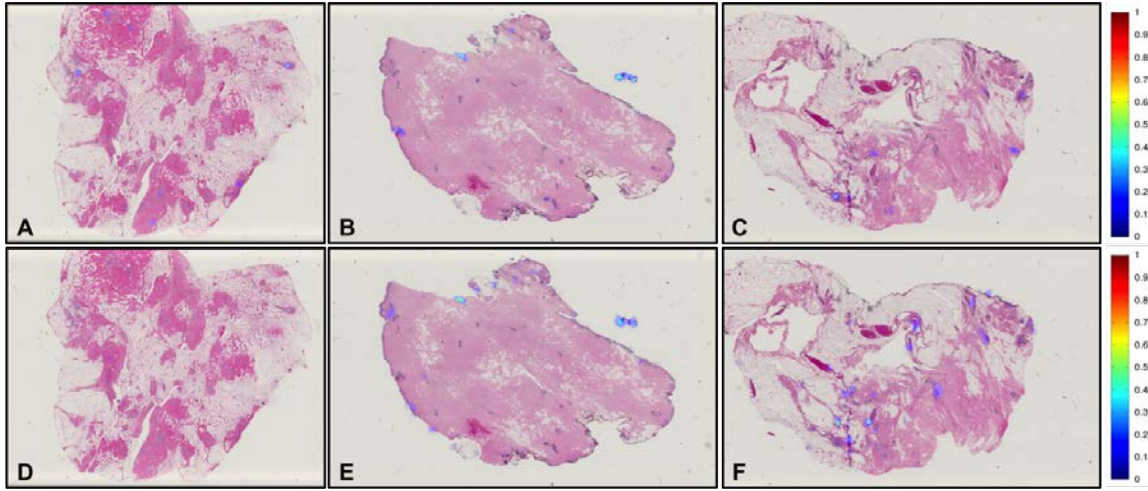


Figure 4-19: The probability maps obtained via the  $\text{ConvNet}_{\text{UHCMC/CWRU}}$  and  $\text{ConvNet}_{\text{HUP}}$  classifiers on whole-slide images of normal breast sections from the UHCMC/CWRU and NC data cohorts are shown in panels (A-C) and (D-F) respectively.

training set employed for the ConvNet model. Exemplars of DCIS and LCIS could, in future work, be included as part of an expanded learning set, as it would not doubt improve the classification performance and generalizability of the model. Additionally and as part of future work, the learning set could be expanded to include other rare variants of invasive ductal carcinoma, such as mucinous invasive carcinomas.

Batch effects are one of the main sources of variation in evaluating the performance of automated machine learning approaches. These batch effects include stain variability due to different histology protocols from different pathology labs and variations in the digitization process on account of the use of different slide scanners [90]. Our results suggest a slight batch effect when we train our ConvNet with two different data cohorts. However, the use of all available training data (HUP and UHCMC/CWRU) results in a more confident, accurate and robust ConvNet classifier. Clearly, increasing the training dataset size and diversity results in a better and more robust algorithm. ConvNet also performs better when a case has only a single morphologic pattern of invasive breast cancer in the whole slide images. Cases with a mixture of invasive and in situ breast cancer resulted in a reduction in the overall accuracy of the ConvNet classifier (in situ tumors may be incorrectly classified as invasive carcinoma). One way of potentially reducing batch effects is to apply color normalization on the digitized images prior to training or application of the ConvNet classifier. To reduce false positive classification errors we are exploring the expansion of the current two class ConvNet classifier into



a multiclass predictor. This will allow for the ConvNet classifier to explicitly deal with the detection of additional subtypes of invasive and in situ breast cancers. One interesting aspect of our work is that the trained ConvNet classifier can be easily integrated into other computational frameworks such as automated tumor grading of ER+ breast cancer subtypes in histopathology images [10]. Our automated invasive cancer detection algorithm could thus pave the way for creation of decision support tools for breast cancer diagnosis, prognosis and theragnosis for use by the pathology community. Future studies will address these opportunities.

### **4.2.5 Conclusions**

We presented an automatic invasive breast cancer detection method for whole slide histopathology images. Our study is unique in that it involved several hundred studies from multiple different sites for training the model. Independent testing of the model on multi-site data revealed that the model was both accurate and robust. This method can be applied to large, digitized whole slide images to detect invasive tissue regions, which could be integrated with other computerized solutions in digital pathology such as tumor grading.

## **Part III**

### **Accurate and high-throughput whole-slide histopathology image analysis**

# 5 Scaling up convolutional networks through high-throughput adaptive sampling: Applications in Digital Pathology

*For real world application of digital pathology it is require efficient and accurate methods to analyze the digitized large whole-slide images. This chapter presents a novel adaptive sampling method for high-throughput whole-slide image analysis applied to the problem of automatic invasive breast cancer detection (see into Section 5.1). The following section correspond to a paper presenting the details of the method and its application.*

## 5.1 High-throughput whole-slide image analysis in Invasive Breast Cancer

Precise invasive tumor region detection on whole-slide images is the very first step for a pathology diagnostic workflow and allows further analysis of tumor differentiation using standard grading schemes. Convolutional neural networks (CNNs) are the most popular representation learning methods for computer vision tasks involving image classification and object detection. While some researchers have employed CNNs for problems in digital pathology including tumor and mitosis detection, these approaches typically use relatively small image sizes (200x200 pixels). Application of CNNs directly to learning patterns from whole slide images is not feasible because the overall size of a CNN depends on the size of the input image; for a WSI this would translate to trillions of parameters in the CNN. To alleviate this issue, in this paper we present a novel method, High-throughput Adaptive Sampling for whole-slide Histopathology Image analysis (HASHI), which involves: i) a new time-efficient adaptive sampling method based on probability gradient and quasi-Monte Carlo sampling, and, ii) a state-of-the-art representation learning classifier based on CNNs. In this work we applied HASHI to the problem of automated detection and quantification of invasive breast cancer extent

on whole-slide images. The method was trained and validated using three different data cohorts involving near 500 cases and then independently validated on an additional 195 studies from The Cancer Genome Atlas (TCGA). The experimental results show that HASHI (1) outperforms state-of-the-art classification approaches involving hand-crafted features, (2) the adaptive sampling method is an effective strategy to deal with large whole-slide images without compromising prediction accuracy, and (3) on an independent test dataset, our approach is effective and robust, with the prediction results being highly consistent across data from multiple sites, scanners, and platforms. *The complete content of this chapter has been submitted as a research article in the Journal of IEEE Transactions on Medical Imaging (see [34]).*

### 5.1.1 Introduction

The advent of whole-slide digital scanners has allowed for rapid digitization of histopathology slides, making these digitized slides images easy to store, visualize, share and analyze using computational tools. This rapidly growing field of Digital Pathology [113] is resulting in one of the newest forms of “big data”. Whole-slide images (WSI) in histopathology are large, typically each WSI could have a spatial resolution of  $80,000 \times 80,000$  pixels and approximately 20 GB in storage size. Additionally, projects like the The Cancer Genome Atlas (TCGA) [126] have resulted in the creation of very large digital slide repositories. The TCGA currently hosts 11,079 cancer studies involving 34 different types of cancer and hosting over 1,095 Terabytes (1 Petabyte) [126].

This high volume of data requires the development and application of high throughput computational image analysis approaches for mining the digital image data. In particular, representation learning and deep learning approaches have been shown to be promising for automated interpretation and analysis of large volumes of images. These approaches have shown significant better performance compared to several state-of-the-art computer vision approaches for tasks such as object detection, object recognition and image annotation [16, 145]. Deep representation learning refers to a family of machine learning methods which learn multiple levels of representation to model complex relations among data. These methods attempt to discover more abstract features via higher levels of representation which then could help facilitate high-level decision tasks such as classification or prediction [15, 16, 47]. In the particular case of image analysis, convolutional neural networks (CNN) represent one of the most successful deep representation learning methods. CNNs are multilayer neural networks, combining different types of layers (convolutional, pooling, classification) which then need to be trained in a supervised manner [16] for image analysis and classification tasks.

Most previous approaches employing CNNs for image classification and object detection

have focused on very small images (e.g.  $224 \times 224$  pixels) [56, 77, 93]. Some of these previous works have involved the application of CNN to histopathology image analysis [28, 32, 36, 116, 121, 159]. However, all of these approaches have limited their analysis to small regions of interest within the larger WSI. The main reason is that the overall size of the network depends on the size of the input image. For instance, a CNN with an input image of  $200 \times 200$  pixels and 250 feature maps in the first convolutional layer would involve 10 million hidden units, whilst the same architecture with an input RGB color image of size  $80,000 \times 80,000$  (a typical whole slide digitized image) would require around 4.8 trillion hidden units, which far exceeds the computational capabilities of most current high performance computing clusters by several orders of magnitude. This means that a direct application of the traditional CNN approach for object detection or pixel-level classification in WSIs is not tenable.

Perhaps for this reason very few works involving histopathology image analysis have been attempted on WSIs, most of these focusing on image registration, preprocessing and information systems [25, 45, 123]. A very limited number of approaches have been proposed for tissue classification, tumor detection or grade scoring on WSIs [81, 91, 92, 147].

Breast cancer (BCa) is the most common type of cancer in women and the second cause of death in developed countries [125, 148]. Invasive BCa refers to those breast cancers that have spread from the original site and typically tend to have poorer prognosis [48, 53]. Precise invasive tumor delineation on the pathology slide is the very first step as part of diagnostic workflow for breast cancer. This enables the subsequent analysis of tumor differentiation by the pathologist, via the Bloom-Richardson and Nottingham grading schemes [62, 71]. While approaches for breast cancer grading have been previously presented [10], these approaches need for the target region of interest to first be defined.

This paper presents a novel, accurate and high-throughput framework that combines the powerful capabilities of CNN models for image recognition and an adaptive sampling method for high-throughput analysis to detect invasive breast cancer on whole slide images. The method is based on a CNN tile classifier which is used to estimate the probability of the presence of invasive BCa in a small WSI region. Instead of applying the tile classifier densely over the entire WSI, the method adaptively chooses WSI regions with high uncertainty of the tissue type, either invasive or not. These regions are typically associated to the borders of tumor, which exhibit mixtures of tissues. The rationale behind the approach is that regions where the predictor has a greater uncertainty about the type of tissue the region represents, will require more tile samples for improving the confidence of the prediction model. Conversely those regions where the predictor has higher certainty about the tissue class may require fewer tile

samples for making a good prediction. Regions that are identified by the prediction model as having a high certainty of belonging to a specific tissue class (invasive or not) typically tend to present similar visual characteristics in its micro-environment neighbourhood associated to their morphological and architectural histopathology features. By contrast, regions that represent a mixture of tissue classes tend to result in a low certainty of class belongingness, typically at the boundaries between two or more tissue interfaces. This information could be exploited to determine those regions where more samples are needed, however each type of tissue could be located in arbitrary sections of the WSI and could have different sizes and shapes, and this is not known in advanced. To address the problem of identifying where to selectively sample regions corresponding to a specific class, we present a new sampling strategy that alternates between exploration and exploitation. The initial exploration involves a random sampling in turn providing a coarse overview of the tissue type distribution in the WSI. This information is then used to perform detailed spatial analysis of small regions, or local exploitation, via a denser sampling of local regions with higher uncertainty. This process is iterated several times.

More specifically our approach integrates: i) a new high-throughput adaptive sampling method based on quasi-Monte Carlo sampling, and ii) a state-of-the-art representation learning classifier based on CNN. In this work we applied our new approach to the problem of automated detection and quantification of invasive breast cancer extent on whole-slide images. The method was trained and validated using three different data cohorts involving near 500 cases and then independently validated on an additional 195 studies cohort from TCGA.

### 5.1.2 Previous and related work

A number of histopathology image analysis methods focused on feature extraction have been recently presented, where the goal has been to identify the ability of image features in conjunction with a machine learning classifier to predict presence or severity of disease from surgical or biopsy tissue specimens [6, 10, 68, 73, 105, 106, 108, 150].

Most approaches involving feature extraction from digital pathology images are based off hand-crafted feature design. These hand-crafted features aim to capture different tissue morphologic and spatial properties including nuclear morphology, nuclear architecture, color intensities, and tissue texture. Table **5-2** details a set of state-of-the-art hand-crafted features in histopathology image analysis and breast cancer digital pathology tasks [10, 155, 159]. These features are used as baseline in this work.

Despite that there are a lot of work based on hand-crafted features taking advantage of knowledge domain priors and assumptions, there are some limitations. For instance,

these features are more sensible to visual variability and artifacts coming from different sources and processes such as staining, magnification, digitization, etc. [6]. In addition, these features are approximations based on mathematical and statistical formulations to represent the visual content that does not capture all the relevant characteristics and complex mixtures and relationships of the histopathology patterns [40]. In contrast, representation learning automatically learns the transformation of data that make easier the extraction of useful information for posterior high-level tasks using one level or multiple levels of representation (i.e. deep learning) [16, 93, 145]. The main difference in comparison to hand-crafted features is that representation learning learns the more appropriate representation or features without any prior domain knowledge, learning the representation directly from the data.

Recently, approaches based on representation learning and deep learning have been applied for histopathology image analysis, either in a supervised or unsupervised manner. Most of previous studies are based on supervised learning (e.g. tumor and mitosis detection [28, 32, 36, 116]) and relatively few approaches are geared towards unsupervised learning [5, 32, 33]. In fact, the more successful representation learning approach in histopathology image analysis had been convolutional neural networks, outperforming hand-crafted features in several problems [156].

Kothari et al. [92] provide an excellent review of the state of the art in histopathological whole-slide imaging informatic methods, associated challenges, and future opportunities. They describe how most approaches to feature analysis of whole slide images are typically limited to manually selected regions of interest (ROI). In [147], the authors describe a multi-resolution framework for tile-based tissue classification to determine the grade of neuroblastic differentiation. Kothari et al. [91] proposed a visualization framework for studying visual morphological patterns across 1,301 histopathological whole-slide images from 571 patients with ovarian serous cystadenocarcinoma from TCGA. In [81], the authors assessed the impact of different classification algorithms and features sets on both accuracy and computing time for quantification of necrosis in whole-slide images.

Huang et al. [83] attempted to address the problem of time-efficient analysis of breast cancer whole-slide images to determine the nuclear pleomorphism score. They used sparse coding to learn in an unsupervised way the visual representation combined with first- and second-order statistics of multivariate Gaussian distributions. These statistics were then employed in conjunction with a machine learning classifier (support vector machine) to identify invasive and non-invasive cancer patches over whole slides at a low magnification. ROIs are then selected from higher nuclear pleomorphism score regions using a dynamic sampling based on Voronoi tessellation. The final nuclear pleomorphism scoring is calculated in detail from the high-scaled versions of the ROIs

selected.

In this paper we present a new approach, combining convolutional neural networks and a novel adaptive sampling based on probability gradient and quasi-Monte Carlo sampling, for the task of invasive breast cancer detection from whole-slide images. Monte Carlo and Quasi-Monte Carlo methods have been traditionally applied for simulation and sampling of complex probability density functions [144]. While these methods have been previously applied in the medical imaging domain for problems involving radiation dose calculation [27, 88] or for segmentation of different biological structures and organs from medical imaging data [3, 66, 85], to the best of our knowledge, the new method presented in this paper represents the first attempt to employ Quasi-Monte Carlo sampling for efficient whole slide histopathology image analysis.

The main limitations of these previous approaches has been that the analysis has been limited to small ROIs within the larger whole slide images. Additionally, hand-crafted features tend to be very specific to particular domains or data sources and not seamlessly generalizable to different tasks or applications. Finally most of these approaches have involved evaluating the methods on a relatively small cohort of cases typically originated from a single institution. Consequently it is not clear whether these approaches will actually be useful for routine clinical practice.

Unlike previous approaches [28, 32, 36, 81, 83, 91, 116, 147], our method has the following advantages and makes the following contributions: i) accurate and reproducible detection of invasive breast cancer regions on new unseen whole slide images, ii) ability to generalize to images acquired from different data sources and domains since it is based on the representation learning method of CNNs, and iii) a new high-throughput adaptive sampling method that makes our approach feasible for whole slide images and is an order of magnitude more efficient compared to the naive implementation of CNNs, while simultaneously not compromising detection accuracy.

In order to explicitly address the issues of variability in staining, slide preparation, and scanning across multiple sites, our training and validation sets were comprised of slide images from multiple different institutions. In contrasts with works such as Huang et al. [83], where a total of nine WSIs from a single site were employed for training and testing of a classifier for cancer detection, in this work we employed almost 600 patient slides from four different sites for training, validation, and independent testing of our CNN model.

### 5.1.3 Methodology

Figure 5-1 presents the general overview of the high-throughput framework for invasive BCa detection in WSI. Training exemplars for the CNN are generated by pathologists



who annotated the regions of invasive BCa on WSIs. The training phase of the CNN uses as input a tile-based dataset obtained by applying a regular sampling of WSIs from the training data cohort to extract tiles of a fixed square size both from within pathologist annotated invasive and non-invasive tissue regions. The prediction stage on new unseen WSIs involves the following steps: first, tiles are extracted from the WSI using random sampling; the CNN classifier is applied to each tile; the prediction produced by the CNN is used to build an interpolated probability map which is then used to identify those regions with high uncertain about the tissue type, whereby is needed a more dense sampling, this is done choosing the high gradient magnitudes associated to border of tumors or mixture of tissue types; the newly sampled exemplars are used to produce an improved probability map estimation; the process is iterated several times, producing as result a final invasive BCa probability map. The details of each step are explained in the following subsections.

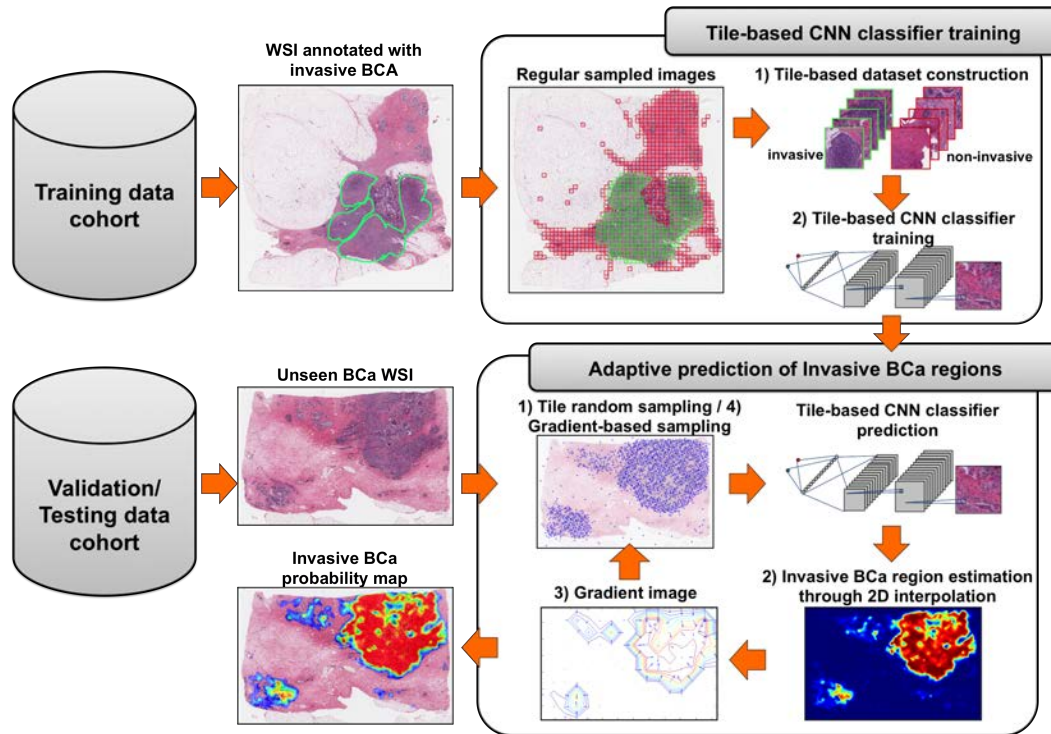


Figure 5-1: Overview of the high-throughput adaptive sampling for whole-slide histopathology images method (HASHI) based on convolutional neural networks (CNN) for automated detection of invasive BCa in WSIs.

### Adaptive gradient-based sampling for efficient Invasive BCa prediction in WSI

Algorithm 3 describes the adaptive gradient-based sampling strategy, which iteratively refines an initial coarse estimation of an invasive BCa probability map. Inputs to the algorithm include a WSI  $X$ , the algorithm parameters: maximum iterations  $T$  and number of sample points per iteration  $N$ . The algorithm begins with a tile sampling process resulting in  $N$  tile samples. Each tile is classified using the CNN-trained model  $M$  to obtain the probability of the presence of invasive BCa at the particular location of each tile. By interpolating the probabilities calculated at each tile sample, a probability map  $P$  for the WSI is obtained. In order to determine regions with higher uncertainty, the gradient of the probability map is calculated  $G$ . The values of the gradient of the probability map are used to prioritize the sampling selection of new tiles for the next iteration. The process is repeated until the maximum number of iterations  $T$  is reached.

---

**Algorithm 3** Adaptive gradient-based Quasi-Monte Carlo sampling.

---

**INPUT:**

$M$ : CNN-trained model  
 $X$ : whole-slide image  
 $T$ : maximum iterations  
 $N$ : number of samples per iteration

```

1:  $samples \leftarrow$  uniform sampling ( $X, N$ )
2: for  $i = 1$  to  $T$  do
3:    $predictions \leftarrow$  tile classification ( $M, samples$ )
4:    $P \leftarrow$  invasive BCa probability map interpolation ( $predictions, samples$ )
5:    $G \leftarrow$  probability gradient ( $P$ )
6:    $samples \leftarrow$  gradient based sampling ( $G, X, N$ )
7: end for
   return invasive BCa probability map  $P$ 

```

---

### Tile-based CNN classifier training

**Tile-based dataset construction and preprocessing** To extract tissue samples for the training process of the tile-based classifier, a regular sampling was performed on each WSI from the training set. Only tiles corresponding to tissue regions were included, background regions were ignored. Additionally a tile sample was considered to be a positive sample (i.e. invasive BCa) if a certain pre-defined proportion of its

area overlaps with the region manually annotated by pathologists as invasive tumor, otherwise it is labeled as a negative sample (i.e. non-invasive BCa) [36].

Each image patch or tile is converted from RGB to YUV color space and normalized to a mean of zero and variance of one. This initial step allows to decorrelate and accentuate the differences between the input image tiles and helps to accelerate gradient-based learning [101].

**Tile-based CNN classifier** We evaluated different CNN architectures (this is discussed in detail in Section 5.1.4), the best one being a 2-layer CNN (CS256-FC256), illustrated in Figure 5-2. This architecture is composed of a convolutional and a pooling layer of 256 units followed by a fully-connected layer of 256 units. The classification layer is a softmax classifier with two outputs (invasive and non-invasive) activated by a logistic regression function. The convolution layer involves application of a 2D convolution of the input image with a kernel of  $8 \times 8$  pixels. The pooling (or subsampling) layer applies a spatial L2-pooling function without overlapping using a pooling kernel of  $2 \times 2$  pixels for each feature map obtained from the convolution step. An advantage of the L2-pooling function is that it allows the learning of invariant features [97]. The output of the pooling layer is fed to a fully-connected layer followed by a final classification layer. The training process uses the set of tiles sampled from both the invasive and non-invasive tissue regions. The CNN model is then trained using a stochastic gradient descent approach [20] to minimize a softmax loss function, which is presented in Eq 5-1:

$$L(w) = -\log \left[ \frac{e^{x_i(w)}}{\sum_j e^{x_j(w)}} \right] \quad (5-1)$$

where  $w$  represent the weights of the network,  $x_i(w)$  correspond to the output of the  $i$ -th neuron in the fully-connected layer, preceding the output layer. The CNN training process involves searching for a weight vector  $w$  which aims to minimize this loss function (Eq. 5-1). During testing, the probability that a particular image tile corresponds to invasive BCa is calculated by applying an exponential function to the output of the CNN. The implementation of the CNN model, its training and testing were performed using Torch 7, a scientific computing framework for machine learning [29].

### Adaptive prediction of invasive BCa region

In order to apply the classifier for predicting the likelihood of individual tiles within the WSI representing invasive cancer or not, the classifier would need to be repeatedly

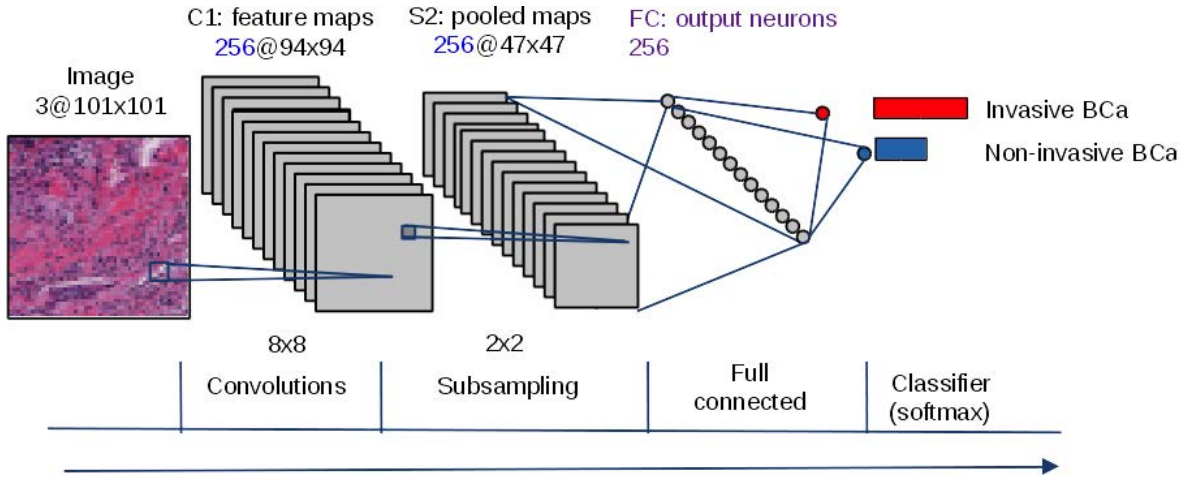


Figure 5-2: Illustration of the Convolutional Neural Network architecture used to distinguish between invasive BCa and non-invasive BCa. Amongst the various architectures considered, this architecture was found to perform the best.

applied to each tile in the WSI. For a WSI of size  $80,000 \times 80,000$  pixels, a tile sampling involving patch sizes of  $101 \times 101$  pixels translates to over  $6.39 \times 10^9$  predictions, which is clearly infeasible. Hence our approach is to perform the prediction on a sample of patches from the WSI and extrapolate this prediction to the whole image. Traditional ways of performing this sampling are: dense, regular and random with a uniform distribution [128]. The method presented in this paper uses an adaptive scheme which performs a guided sampling that focuses on those image areas with higher uncertainty. Each of the individual steps involved in the adaptive tile based classification are described below.

**Tile sampling** The goal of this step is to select a set of tiles (in this case  $101 \times 101$  tiles) from the WSI, which will be later used to create a probability map over the entire WSI. The tiles may be selected by deterministic (dense or regular) or random sampling. The different sampling approaches are discussed here below.

**Regular sampling:** This strategy involves sampling tiles at equally spaced intervals on a regular grid. For instance, given a WSI of  $K \times K$  size and tiles of  $k \times k$  size, the step size  $s$ , in both X and Y axis is  $1 \leq s \leq k$ . The extreme case is using a step size  $s = 1$ , which means an expected number of samples of  $(K - k)^2$ . This case corresponds to a dense sampling of the WSI.

**Uniform random sampling:** Regular sampling is deterministic. An obvious alternative strategy is random sampling, i.e. to select the tiles using random coordinates

generated from a particular probability distribution. Without *a priori* knowledge of the image content, a uniform probability distribution is a natural choice for the random sampling algorithm.

**Quasi Monte Carlo sampling:** Uniform random sampling may concentrate samples in some regions of the image while leaving other regions under-represented. This may not be the most efficient strategy since the predictions on overrepresented regions will be redundant.. Regular sampling may also produce sub-optimal results if the boundaries of the region to be estimated is not aligned with the axis. Quasi Monte Carlo (QMC) sampling represents a good compromise between regular and random sampling. QMC sampling accomplishes this by using a random sampling procedure based on deterministic (pseudo-random) sequences designed to have low discrepancy, where discrepancy is a measure of the uniformity of a distribution of finite point sets [127]. This property is an advantage for QMC in contrast to Monte Carlo methods (based on random sampling) since QMC does not result in clumping, which in turns results in better accuracy for the sampling process [22]. We chose the Sobol and Halton sequences [122, 127, 152] for our iterative adaptive sampling method. With these sequences it is possible to incrementally add sample points without discarding those already previously generated [129]. Thus, for a 2D image we can generate  $N$  points per iteration for each low discrepancy sequence. Figure 5-3 depicts an example of sampling points generated using random, Sobol and Halton sequences, respectively.

**Invasive BCa probability map estimation** The sampled tiles are fed to the CNN classifier to determine the probability of the presence of invasive or non-invasive BCa in each particular tile. Cubic interpolation is then applied to extrapolate this estimation to all the pixels in the WSI, resulting in an invasive BCa probability map at the end of each iteration.

**Probability gradient** A gradient image  $\nabla P$  is calculated to identify the directional changes of the probability map  $P$  as follows:

$$\nabla P = \frac{\partial P}{\partial x} \hat{x} + \frac{\partial P}{\partial y} \hat{y} \quad (5-2)$$

where  $\frac{\partial P}{\partial x}$  is the gradient in the  $X$  direction and  $\frac{\partial P}{\partial y}$  is the gradient in the  $Y$  direction. Then, the gradient magnitude image  $|\nabla P|$  is calculated to identify regions with high or low variations among tissue types in the probability map  $P$ . High values correspond to heterogeneous spatial distribution of both invasive and non-invasive tissue types in tumor boundaries (strong changes), whereas low values corresponds to homogeneous

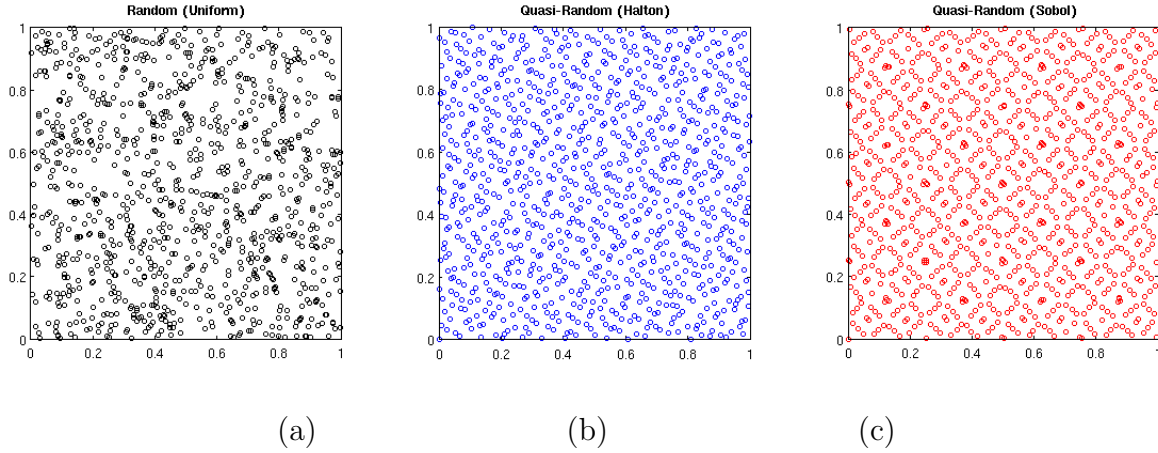


Figure 5-3: Example of 1000 2D-points generated by (a) random (uniform distribution) and quasi-random sampling strategies, (b) Halton and (c) Sobol sequences. (a) We can notice that random generation results in clumping and there are more regions without samples. On the other hand, (b) Halton and (c) Sobol sequences have a more uniform distribution over the whole space with very few clumps, i.e. low discrepancy.

distribution of either invasive or non-invasive tissue types (soft changes). Formally, the gradient magnitude image  $|\nabla P|$  is defined as follows:

$$|\nabla P| = \sqrt{\left(\frac{\partial P}{\partial x}\right)^2 + \left(\frac{\partial P}{\partial y}\right)^2} \quad (5-3)$$

where the magnitude  $|\nabla P|$  has low values or zero if the local spatial regions of  $P$  has similar or constant values. In contrast, the magnitude  $|\nabla P|$  has high values if the local spatial regions of  $P$  has different values changing from low towards high probability and vice versa.

**Gradient-based sampling selection** The probability gradient enables a more intelligent spatial sampling of points, with a more dense sampling directed at high magnitude regions within the gradient map (representing transitional areas from one tissue type to another where the uncertainty is higher), while fewer samples are extracted from homogeneous regions (representing a single tissue type region with smaller uncertainty). This is accomplished by the following procedure: first,  $2N$  samples are generated using the corresponding sampling mechanism (random or QMC); second, the samples

are ranked according to the magnitude of the gradient in the corresponding position; finally, the top  $N$  samples are returned.

### 5.1.4 Experimental design

#### Breast Cancer data cohorts

Table 5-1: Breast cancer data cohorts used for training, validation and testing in the experimental evaluation.

ID	Site	Number of patients	Scanner	Dataset	Use
HUP	Hospital of the University of Pennsylvania	239	Aperio	Training	Experiment 1, Experiment 2
UHCMC/CWRU	University Hospitals Case Medical Center/Case Western Reserve University	110	Ventana	Training	Experiment 3
HUP + UHCMC/CWRU	Hospital of the University of Pennsylvania and University Hospitals Case Medical Center/Case Western Reserve University	349	Aperio and Ventana	Training	Experiment 3
CINJ	Cancer Institute of New Jersey	40	Aperio	Validation	Experiment 1, Experiment 2
TCGA	The Cancer Genome Atlas <sup>1</sup>	195	Aperio	Testing	Experiment 3

The data used in this study are H&E-stained histological slides from patients with estrogen receptor-positive (ER+) breast cancer obtained from four different sites. The WSIs were digitized with Aperio or Ventana scanners. We used only those images

which had been scanned in at a 40x magnification. Images were further downsampled by a factor of 32:1 for each axis. As illustrated in Table 5-1, different data cohorts were used for training, validation or testing. The training and validation sets were used for model parameter tuning and optimization. Independent model evaluation was performed on the independent test set.

Two expert pathologists (NS, HG) independently provided the ground truth annotations of invasive breast cancer regions over digitized whole slide images for each data cohort (NS for HUP and CINJ; HG for UHCCMC/CWRU and TCGA). The pathologists manually delineated the invasive regions at 2x magnification using the viewing software applications ImageScope v11.2 from Aperio and Ventana Image Viewer v3.1.4 from Ventana.

### **Experiment 1: Invasive BCa tile classification performance using CNN vs handcrafted features**

Our goal in this experiment was to compare the most commonly used hand-crafted features in histopathology image analysis, in general and for breast cancer diagnosis, (Table 5-2) [6,73,106,155,158], against different CNN based feature learning architectures in the task of tile-based tissue classification between invasive and non-invasive BCa. This experiment uses as training data set the HUP data cohort (239 cases) and as test dataset the CINJ data cohort (40 cases). Parameter tuning was performed using cross validation over the training dataset. The performance of the classifier was evaluated using the area under the ROC curve (AUC).

Each hand-crafted feature listed in Table 5-2 was combined with each of two classifiers: random forests (RF) and support vector machines (SVM). For RF, the training step was used to optimize the parameter corresponding to the number of trees, while for the SVM different kernel functions were evaluated: linear, radial basis function (RBF), intersection, Chi-square ( $\chi^2$ ), and Jenson-Shannon's.

For the CNN-based approach, three different architectures were evaluated. The first was the architecture employed in [36] which was a 3-layer CNN, called ConvNet. This architecture comprises of 16 neurons in the initial convolutional and pooling layers, 32 neurons in the second stage, and 128 neurons in the third fully-connected layer (CS16-CS32-FC128). The second architecture explored was the one that was previously successfully applied to the problem of mitosis detection in breast cancer histopathology images [28], which comprises four layers of convolutional and pooling neurons with 16 neurons in each, and a fully-connected layer of 128 neurons (CS16-CS16-CS16-CS16-FC128). The third architecture explored was a 2-layer CNN with 256 neurons in the first layer and 256 neurons in the fully-connected layer (CS256-FC256).

In order to determine the statistical significance of the performance difference among



methods, we applied a multiple comparison Kruskal-Wallis test using the following procedure: we built 100 different datasets with the 60% of the instances from HUP data cohort applying bootstrap sampling. The AUC for each method in each dataset was evaluated and the methods were ranked according to performance for each group. Based on these ranks, the Kruskal-Wallis test statistic was calculated and a post-hoc Tukey’s honestly significant difference criterion was applied to check for pairwise differences between methods.

### **Experiment 2: Evaluation of the impact of the sampling strategy on the effectivity and efficiency of the method**

Seven different sampling methods were evaluated to determine the more efficient strategy for WSI analysis in terms of both detection accuracy and computing time. The baseline sampling method is (i) regular sampling (*regular*) which takes equally-spaced samples by varying the step size. For the random and pseudo-random sampling methods, we evaluated (ii) uniform random sampling (*uniform*), (iii) quasi-Monte-Carlo sampling using the Sobol sequence (*qmc-sobol*), and (iv) quasi-Monte-Carlo sampling using the Halton sequence (*qmc-halton*). In addition, sampling strategies including information from the gradient image to select more samples on high uncertain regions per iteration were combined with the previous sampling strategies, (v) gradient-based uniform sampling (*grad-uniform*) and (vi) gradient-based quasi-Monte-Carlo sampling, using either Sobol (*grad-qmc-sobol*) and (vii) Halton (*grad-qmc-halton*) sequences. All sampling approaches (with and without incorporation of gradient image information) were applied iteratively with the same set of parameters: 20 iterations and 100 samples per iteration, resulting in 2000 samples for each sampling approach.

This experiment used the best performing CNN model identified in Experiment 1 in conjunction with all the various sampling strategies that were evaluated. The performance of the classifier in conjunction with the different sampling strategies was evaluated on the CINJ data cohort using 12 cases.

The output of our method is an invasive BCa probability map over the WSI, i.e. a measure of the probability of presence of invasive BCa for each pixel in the WSI. Since the ground-truth annotations provided by pathologists correspond to binary masks indicating the presence of invasive BCa in particular regions, we need to extract a binary mask from the estimated probability map that can be compared against the ground truth. This is accomplished by defining a threshold  $\alpha$  and extracting the region enclosed by the  $\alpha$ -level set of the probability map. A good value of  $\alpha$  is calculated using cross validation over the training dataset.

The quantitative evaluation was done by measuring the Dice coefficient between the predicted mask and the ground truth annotation from the expert pathologist. The

Dice coefficient is defined as follows:

$$Dice = \frac{2|P \cap G|}{|P| + |G|} \quad (5-4)$$

where  $P$  corresponds to the predicted binary mask by our method, and  $G$  is the ground truth binary mask from the pathologist annotation. For each of seven sampling strategies, the average Dice coefficient over the test data set is calculated for different number of samples.

### **Experiment 3: Overall method effectivity evaluation over an independent test cohort**

We trained a CNN model with the best configuration found in Experiment 1. Additionally a linear SVM with the best performing handcrafted feature (CF) was trained using the complete training data cohort (i.e. HUP and UHCMC/CWRU,  $n = 349$ ) for both approaches. The testing dataset was a random selection of a subset of ER+ breast cancer whole-slide images from TCGA ( $n = 195$ ). The quantitative evaluation was done by comparing the manual annotations of invasive tumor regions by expert pathologists versus the predictions obtained by each approach in terms of the Dice coefficient measure.

Using the pathologists ground truth mapping of cancer extent, we also identified which regions corresponded to true positive (TP), true negative (TN), false positive (FP) and false negative (FN) errors. The performance measures used to evaluate the accuracy of our method were Dice coefficient (Dice), true positive rate (TPR), true negative rate (TNR), false positive rate (FPR) and false negative rate (FNR).

## **5.1.5 Results and discussion**

### **Experiment 1: Invasive BCa tile classification performance using CNN vs handcrafted features**

Table 5-3 shows the AUC values (mean and standard deviation) across all the images in the independent test set for the three proposed CNN models described in Subsection IV-B (CS16-CS32-FC128, CS16-CS16-CS16-CS16-FC128, CS256-FC256) and the handcrafted features presented in Table 5-2 combined with RF and SVM classifiers. The experimental results in Table 5-3 show that the three CNN classifiers outperform the best combinations of hand-crafted features and classification methods. Additionally, the CNN classifiers exhibit a smaller variance in terms of the AUC measure.

The multicomparison Kruskal-Wallis test, using a post-hoc Tukey’s honestly significant difference criterion, reveals that there is no statistical difference ( $p < 0.05$ ) in terms of critical difference <sup>2</sup> among the CNN classifiers (CS16-CS16-CS16-CS16-FC128, CS256-FC256, CS16-CS32-FC128), and that the two top performing CNN models significantly outperform hand-crafted features. For the rest of the experimental evaluation we used CNN model with the second best performance (CS256-FC256) since it has a simpler architecture (fewer layers).

### Experiment 2: Evaluation of the impact of the sampling strategy on the effectivity and efficiency of the method

Figure 5-4 shows the invasive BCa probability map produced for a test WSI using a common sampling strategy and one configuration of the proposed adaptive sampling, which were presented above: regular sampling (*regular*) and gradient-based quasi-Monte-Carlosampling using Halton sequence (*grad-qmc-halton*). Figure 5-4-A shows a test WSI while Figure 5-4-B shows the ground truth annotation provided by an expert pathologists. Figure 5-4-C is the prediction using regular grid sampling with a step size equal to the tile size (i.e.  $50 \times 50$  pixels). While this sampling is fast (31 secs), the resulting probability maps are extremely coarse and imprecise. Figure 5-4-D shows the probability map obtained using dense regular sampling, which is the extreme case of regular grid sampling where the step size is 1 pixel. The resulting probability map from the dense regular grid sampling is highly specific and detailed. Unfortunately with a run time of 22 hours it is also quite unfeasible for application in a clinical setting. Figures 5-4-E to 5-4-L show the iterative process of the new adaptive sampling method (*grad-qmc-halton*). Figures 5-4-E to 5-4-H visualize the sampled points for CNN classification process from iterations 1st, 2nd, 8th and 20th, respectively. Whilst Figures 5-4-I to 5-4-L depict the invasive BCa predictions using the gradient-based adaptive sampling strategy: *grad-qmc-halton*. Qualitatively, *grad-qmc-halton* sampling strategy shows a comparable performance with dense sampling providing more details than regular sampling. This is corroborated by the quantitative evaluation results shown in Figure 5-5.

The different sampling strategies were quantitatively evaluated in terms of the Dice coefficient for 12 images from CINJ data cohort. The performance of each sampling strategy was evaluated for different number of samples. Figure 5-5 shows the quantitative results of the sampling strategies in terms of Dice coefficient to predict the invasive BCa regions versus the number of tile samples required in the WSI taken per iteration. The x-axis, in a logarithmic scale, corresponds to the number of tile samples required

---

<sup>2</sup>mean differences above the critical difference are suppose to be statistically significant.

by each sampling method and the y-axis corresponds to the Dice coefficient between the region predicted as invasive breast cancer by our approach and the ground-truth annotation provided by pathologists. Note that the number of samples (x-axis) is proportional to the computing time because classifying each tile sample takes the same time. Each sampling strategy is depicted as a line. The regular sampling strategy was evaluated using different step sizes (200, 150, 100, 75, 50, 25, 1 pixels), while the random and pseudo-random sampling strategies (*uniform*, *qmc-sobol*, *qmc-halton*) were evaluated using the same algorithm's parameter configuration, i.e. 20 iterations with 100 samples per iteration.

The experimental results show that adaptive sampling methods (*grad-qmc-sobol*, *grad-qmc-halton* and *grad-uniform*) not only outperform regular sampling and non-adaptive random sampling (*uniform*, *qmc-sobol*, *qmc-halton*) but achieve the same detection performance as dense sampling but with a significantly fewer number of samples having to be classified, resulting therefore in substantial computational time reductions. While dense sampling takes around of 6 millions of tile samples in average and a compute time of around 24 hours per WSI, our adaptive sampling strategies (*grad-qmc-sobol* and *grad-qmc-halton*) achieve a comparable detection performance with 1000-2000 samples taking less than one minute for a WSI.

### Experiment 3: Overall method effectivity evaluation over an independent test cohort

In the third experiment we applied the best combination of tile-based tissue classifier (CS256-FC256) and adaptive sampling method (*grad-qmc-halton*) for our HASHI method ( $HASHI_{grad-qmc-halton}^{CS256-FC256}$ ) to the independent testset from TCGA ( $n = 195$ ). Table 5-4 summarizes and compares the performance of our selected method  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  versus the best classifier based on hand-crafted features, i.e.  $HASHI_{grad-qmc-halton}^{CF-SVM-Linear}$ , in terms of average Dice, TPR, TNR, FPR and FNR. These results show that the CNN-based method  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  has a 75% agreement with the manual annotations from pathologists in terms of Dice, whereas the method using the machine classifier trained with the best hand-crafted feature  $HASHI_{grad-qmc-halton}^{CF-SVM-Linear}$  has a 73% Dice coefficient.

A more detailed analysis of the distribution of the Dice coefficient per case reveals that most cases had a Dice value between 0.7 and 0.9 and an overall median value of 0.8228 using the CNN-based method  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  whereas the best hand-crafted-based features  $HASHI_{grad-qmc-halton}^{CF-SVM-Linear}$  achieved a median value of 0.8007. Some of the cases with the lowest Dice coefficient were because the classifier also identified ductal carcinoma in situ (DCIS), a stage 0 breast cancer that is also sometimes considered as a pre-malignancy. However, since we set a very stringent requirement on only identifying

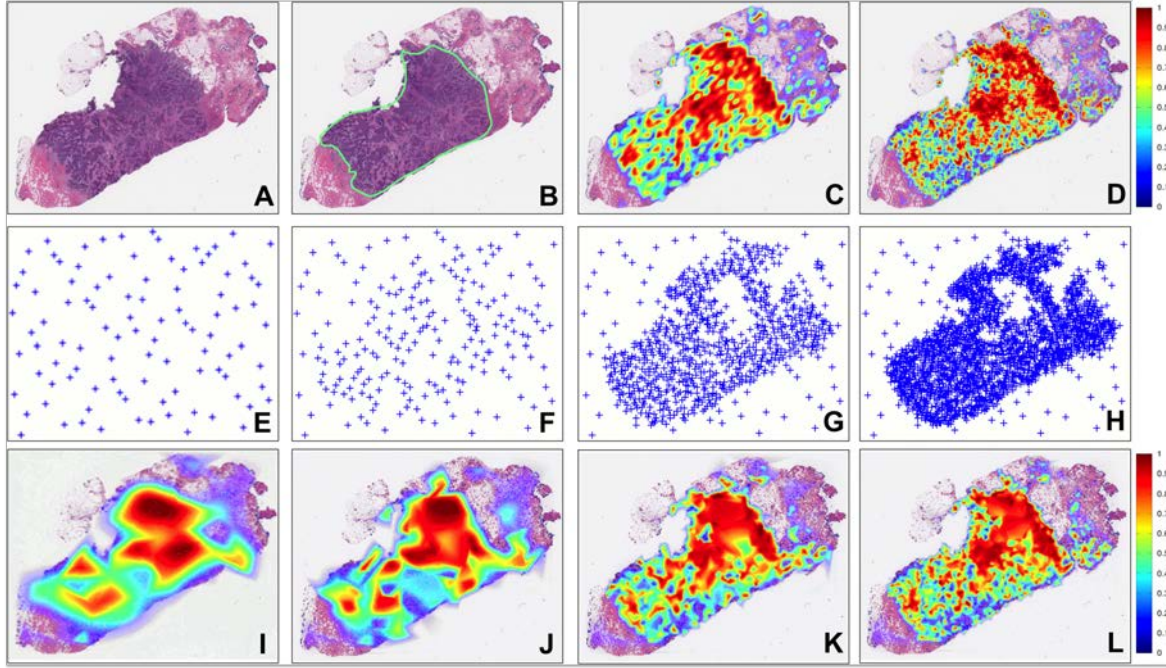


Figure 5-4: Comparison between common sampling methods (*regular* and *dense*) with our novel adaptive sampling method using gradient-based quasi-Monte Carlo sampling (*grad-qmc-halton*). The new unseen WSI (A) with its corresponding ground truth annotation from an expert pathologist (B). The probability maps using regular sampling with step size equal to the patch size (C) and regular dense sampling with step size equal to 1 pixel (D). The new adaptive sampling method shows the iterative process to extract patch samples (E-H) and obtain the probability maps (I-L) for the 1st iteration (E, I), the 2nd iteration (F, J), the 8th iteration (G, K) and the 20th iteration (H, L).

invasive cancer, the detection of DCIS was deemed to be a false positive error. Most other cases with a low Dice coefficient corresponded to slides with poor staining quality.

Figure 5-6 shows the sensitivity of both methods  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  and  $HASHI_{grad-qmc-halton}^{CF-SVM-Linear}$  to the threshold used to determine the tumor area, evaluated in the TCGA test data cohort. Interestingly, the performance of the Dice coefficient of the CNN-based method  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  is more stable and robust achieving high performance, i.e. greater than 0.7, for most of the possible thresholds between 0.1 and 0.6 achieving its optimal value of 0.7586 for a threshold of 0.24. In contrast, the adaptive sampling method using the Linear SVM trained with Color Features, i.e.  $HASHI_{grad-qmc-halton}^{CF-SVM-Linear}$ , is more sensitive to the selected threshold achieving good results only in the interval

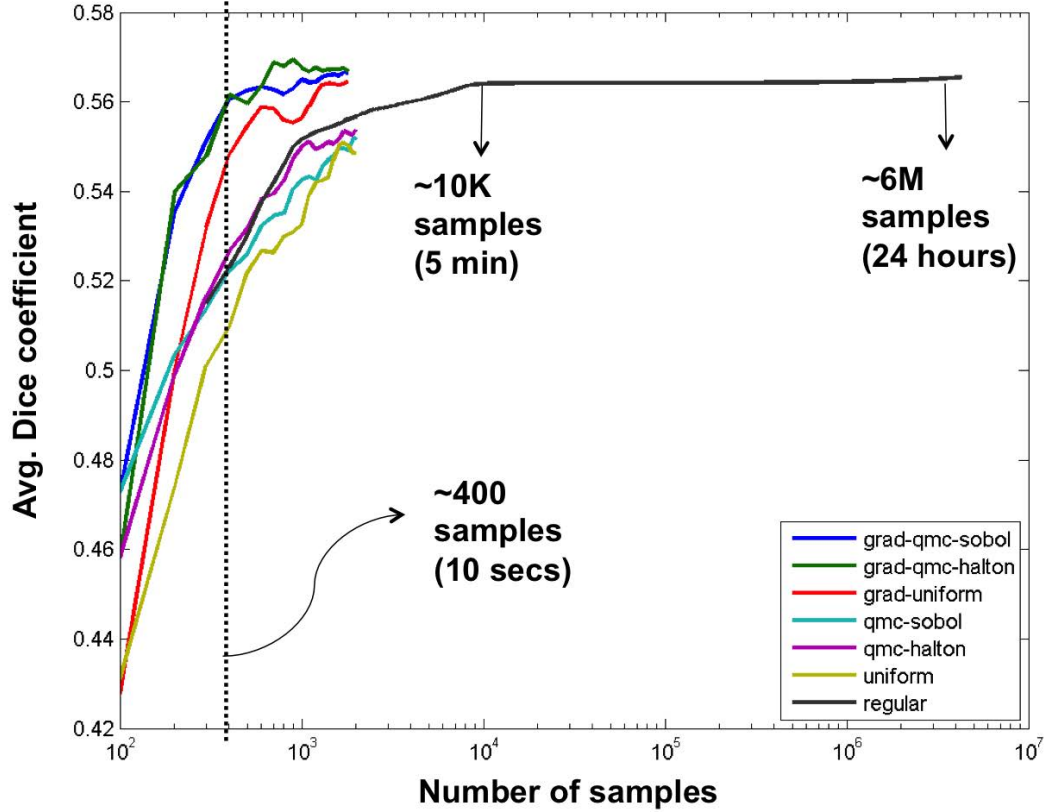


Figure 5-5: Quantitative evaluation in a subset of 12 cases from CINJ data cohort among different sampling strategies for invasive BCa detection on WSIs in terms of the Dice coefficient (y-axis) trained with HUP data cohort versus number of samples (x-axis) required for this prediction using a logarithmic scale.

between 0.35 and 0.45 with its best result of 0.7305 for a threshold of 0.39. These results appear to suggest that  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  is more confident, robust and accurate compared to hand-crafted features for this particular problem.

Figure 5-7 shows the predictions of the HASHI method for representative slides chosen from the test TCGA dataset. Note the good concordance between the predictions from the best configuration of our  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  method (i.e. tile-based tissue classifier of 2-layers CNN and adaptive sampling of *grad-qmc-halton*) and the manual annotations by expert pathologists.

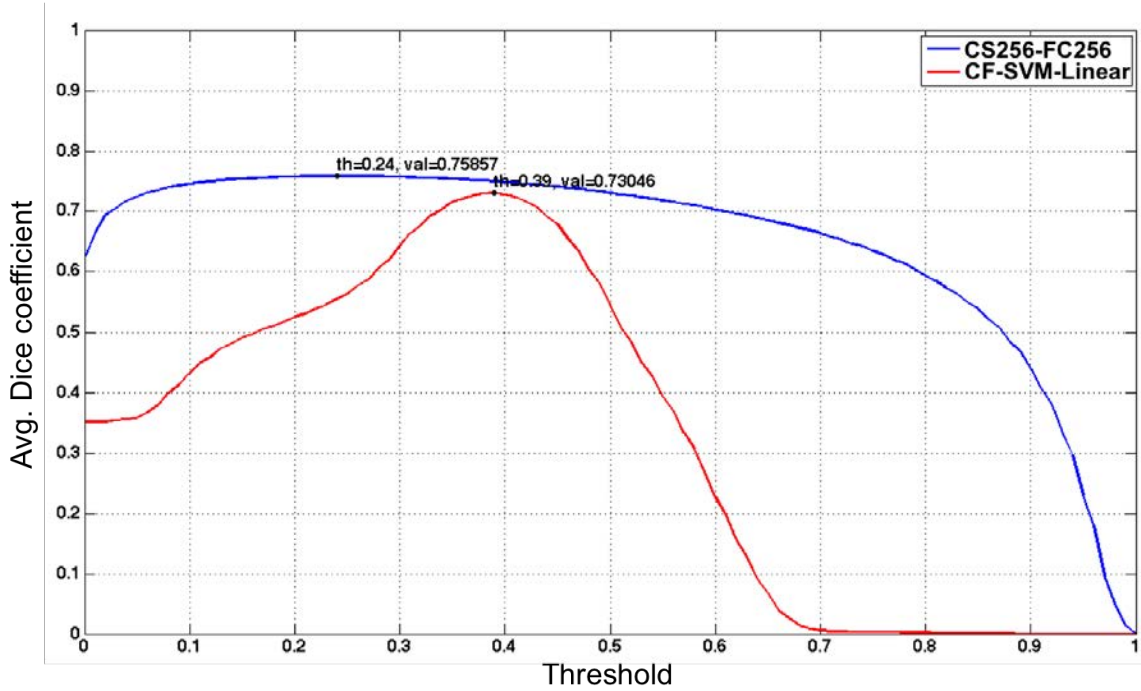


Figure 5-6: Performance comparison between  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  and  $HASHI_{grad-qmc-halton}^{CF-SVM-Linear}$  in terms of Dice coefficient in the independent TCGA test data cohort by varying the classification threshold of the invasive BCa probability map.

### 5.1.6 Conclusions

This paper presented a novel accurate and high-throughput method for automatic invasive breast cancer detection in whole-slide images. We addressed the challenges of high complexity and visual variability of tissues, invasive and non-invasive, and the large size of whole-slide images by combining a state-of-the-art image analysis technique, convolutional neural networks, with an efficient adaptive sampling strategy. The model was trained to learn the most appropriate representation using a considerable amount of whole-slide images from different institutions for training, validation and testing. To deal with large-size images, we developed a novel adaptive sampling method which integrates quasi-Monte-Carlo sampling with a gradient-based adaptive strategy which focuses sampling on those areas with higher uncertainty.

The method was systematically evaluated using nearly 600 whole-slide images from four different institutions for training and validation, and around 200 slides from the publicly available TCGA dataset for testing. The results revealed that our approach is effective and robust, with reproducible results across data from different sources.

The experimental results also demonstrated that convolutional neural network models are more robust and have the best classification performance to discriminate invasive from non-invasive tissue regions in comparison to state-of-the-art features used for histopathology image classification. In addition, the novel adaptive sampling method showed to be an effective strategy to deal with the large size of WSI using less samples than those required by regular dense sampling while keeping the same prediction performance.

Considering that invasive breast cancers are the most common subtypes of breast cancer and that precise delineation is time consuming, but required for further analysis such as tumor grading, the histopathology image analysis framework presented in this paper has a great potential as a support tool that helps pathologists to speed up the invasive tumor identification and localization, alleviating their workload.

Despite of the good results obtained in the evaluation, the method did not perform well on some particular slides which exhibited ductal carcinoma in situ (DCIS) that was confounded with invasive breast cancer. While DCIS is considered as zero stage of breast cancer, it is still not aggressive but sometimes is considered as a pre-malignancy. However, since we set a very stringent requirement on only identifying invasive cancer, the detection of DCIS was deemed to be a false positive error. The other few cases with low prediction performance of invasive BCa were in some WSI corresponding to slides with poor staining quality, which are typically rejected in a common quality control by a pathologist to guarantee an appropriate diagnosis.

These results are a promising and encouraging advance towards translational medicine in digital pathology and cancer research that could be validated and extended, in our future work, with additional cancer subtypes and confounding cases such as atypia hyperplasia, adenosis and ductal carcinoma in situ among others.



Table 5-2: Set of hand-crafted features used for comparison against the CNN based feature learning approach.

ID	Category	Length	Features
CF	Color/intensity	56	Statistics (average, standard deviation, median, and mode) of 14 color channels (blue-ratio, Hematoxylin, Eosin, Red, Green, Blue, Hue, Saturation, Value, Luminance from Lab, a from Lab, b from Lab, U from LUV, V from LUV) [6, 58, 106].
GeF	Geometrical	48	Statistics (average, standard deviation, median, and mode) of geometrical and morphological features (area, eccentricity, equiv diameter, Euler number, Extent, perimeter, solidity, major axis length, minor axis length, elongation, circularity, compactness) [6, 21, 73, 106].
CH	Color Histograms	$8 \times 3$	Histogram of 8 bins for each channel of RGB color space [6, 106].
SH	Shape Index Histogram	$8 \times 3$	Shape index histogram of 8 bins for each channel of RGB color space [57].
MLBP	Multi-scale LBP	$8 \times 3$	Multi-scale local binary patterns histogram of 8 bins for each channel of RGB color space [6].
HF	Haralick features	$26 \times 3$	Statistics (average and standard deviation) of 13 Haralick gray-level [76] concurrence features grabbed at four orientations for each channel of RGB color space [6, 59, 106].
RLF	Run-Length features	$11 \times 3$	Statistics of 11 gray-level run-length matrices properties [70] at four orientations for each channel of RGB color space [73].
GWF	Gabor wavelet features	$71 \times 3$	Statistics (average and standard deviation) of 71 bank of Gabor wavelet filters from eight orientations for each channel of RGB color space [6, 58, 59, 106].
TGF	Topography/Graph features	51	12 features from Voronoi diagram, 8 features from Delaunay triangulation graph, 4 features from minimum spanning tree, and 27 nuclear features [6, 58, 73].

Table 5-3: Classification performance comparison in the CINJ validation dataset among different architectures of CNN models and state-of-the-art hand-crafted features trained with HUP dataset and evaluated in terms of AUC.

Methodology	AUC
CS16-CS16-CS16-CS16-FC128	<b>0.9021</b> +/- <b>0.0097</b>
CS256-FC256	<b>0.9018</b> +/- <b>0.0093</b>
CS16-CS32-FC128	<b>0.8915</b> +/- <b>0.0093</b>
CF + SVM-Linear	0.8711 +/- 0.0947
RLF + SVM-Linear	0.8689 +/- 0.0963
CH+ SVM-Linear	0.8448 +/- 0.1047
SH + SVM-Linear	0.8444 +/- 0.1065
HF + SVM-Linear	0.8385 +/- 0.0942
TGF + SVM-Linear	0.7998 +/- 0.1068
RLF + RF	0.7985 +/- 0.0892
CH + RF	0.7927 +/- 0.0994
GWF + SVM-Linear	0.7911 +/- 0.1210
SH + RF	0.7748 +/- 0.1111
MLBP + SVM-Linear	0.7732 +/- 0.0944
HF + RF	0.7649 +/- 0.1015
TGF + RF	0.7314 +/- 0.1038
GeF + SVM-Linear	0.7291 +/- 0.1414
MLBP + RF	0.7034 +/- 0.0919
MLBP + SVM-INT	0.2134 +/- 0.0931
SH + SVM-RBF	0.1590 +/- 0.1056
CH + SVM-2	0.1308 +/- 0.0918

Table 5-4: Invasive BCa detection performance of our HASHI method on the testing data cohort of TCGA in terms of Dice, PPV, NPV, TPR, TNR, FPR and FNR.

	Dice	PPV	NPV	TPR	TNR	FPR	FNR
$HASHI_{grad-qmc-halton}^{CS256-FC256}$	<b>0.7586</b>	<b>0.7162</b>	<b>0.9677</b>	<b>0.8691</b>	<b>0.9218</b>	<b>0.0782</b>	<b>0.1309</b>
	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	<b>0.2006</b>	<b>0.2204</b>	<b>0.0511</b>	<b>0.1582</b>	<b>0.0764</b>	<b>0.0764</b>	<b>0.1582</b>
$HASHI_{grad-qmc-halton}^{CF-SVM-Linear}$	0.7305	0.6852	0.9658	0.8634	0.9091	0.0909	0.1366
	+/-	+/-	+/-	+/-	+/-	+/-	+/-
	0.2099	0.2378	0.0531	0.1871	0.0762	0.0762	0.1871

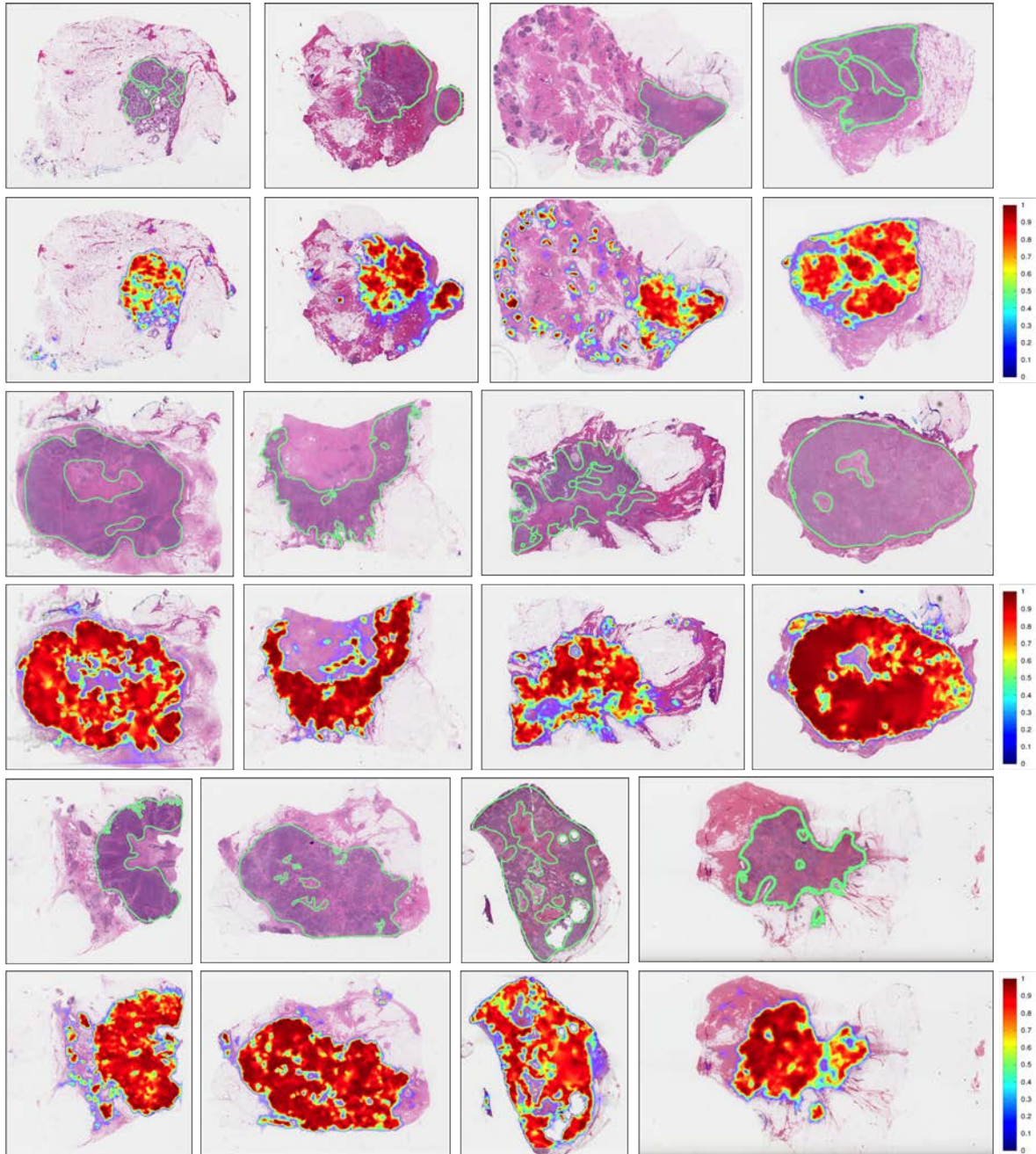


Figure 5-7: Results of the invasive BCa probability maps (second, fourth and sixth rows) predicted by our CNN-based method  $HASHI_{grad-qmc-halton}^{CS256-FC256}$  on representative WSIs from the TCGA test dataset compared to the ground truth annotations from expert pathologists (first, third and fifth rows).

## 6 Conclusions and Perspectives

This thesis has proposed and validated several histopathology image representation learning methods for different digital pathology tasks and cancer types. The aim of these methods is to learn, directly from image collections, the appropriate image representation that is the basis for predictive models with accurate, robust and interpretable results. This thesis has produced novel methods to address several challenging tasks in the digital pathology domain. Some of these methods adapt state-of-the-art computer vision techniques, which were originally designed for natural image analysis, to the challenging particularities of histopathology images. These methods address digital pathology tasks at nuclei, tissue and whole slide levels of analysis. In all cases the proposed methods were able to exploit and learn an appropriate image representation from both small and large histopathology image databases taking advantage of domain knowledge provided by expert pathologists in each cancer type.

The methods were systematically evaluated in diverse digital pathology tasks and cancer types, such as, basal cell carcinoma detection in skin cancer, anaplastic medulloblastoma tumor differentiation in brain cancer, and mitosis and invasive tumor detection in breast cancer.

The representation learning approaches used here start from the assumption that computational learning algorithms can be able to learn not only the high-level tasks but also a good representation of the data for a particular task directly from the database. The main conclusion of this work is that representation learning methods are applicable in digital pathology tasks for different cancer types, improving the results of the previous state-of-the-art approaches. In addition, the new methods introduced more visually interpretable models providing informative support for pathologists. Unsupervised representation learning methods worked better than supervised representation learning methods for small histopathology image databases, while the latter worked better in cases where a large database of annotated cases was available. In contrast with results reported for natural image, we found that the combination of hand-crafted and learned features achieved better performances than each type of feature independently, for histopathology image analysis tasks such as mitosis detection in breast cancer.

An important effort in this thesis was the evaluation of the robustness of the methods

and reproducibility of the results. Because of the scarce availability of publicly available annotated histopathology image databases, it is difficult to find works where the methods are validated on more than one database. In this work, 8 different dataset from different sources were used. In the particular case of invasive BCa detection, three different data cohorts from different institutions, comprising 389 cases, were used for training and parameter tuning, and an additional fourth data cohort, with 195 cases, was used for testing.

A successful translation of research results to clinical practice require efficient methods that can be deployed as part of routine clinical workflows. Time-efficient high-throughput whole-slide image analysis was successfully addressed through a general framework based on supervised representation learning and adaptive sampling achieving fast prediction times, while keeping high accuracy.

All of above contributions and findings suggest that histopathology image representation learning approaches are very promising for real applications in cancer research and clinical decision making to support pathologist's workflows.

## 6.1 Perspectives

This thesis was developed attempting to provide novel computational solutions to support digital pathology workflows. For pathologist's routine is important to have confident, interpretable, quantitative and robust tools for diagnosis support in his/her clinical practice. Successful deployment of these methods could impact different aspects in pathology by alleviating the pathologist's workload for early detection and screening process, providing objective and quantitative image-based analysis for reproducible and robust diagnosis, resulting in more specific diagnosis and guided treatments of patients. The translation of these methods to clinical practice and in-silico cancer research requires further evaluation in additional data cohorts and cancer types.

During the development of this thesis important challenges were addressed for automatic histopathology image analysis in several digital pathology tasks. However, these challenges are far from being totally solved, on the contrary they are the starting point of new directions for future research work:

1. **Interpretable learning models.** Conventional hand-crafted features for image representation and machine learning methods for high-level tasks usually are abstract and with low interpretability. Despite representation learning methods such as bag of features, dictionary learning, sparse coding and deep learning, learn visual features similar to the ones detected by receptive fields of the human visual system [84], still they are not interpretable enough for pathologists.

The open challenge is to provide semantic and visually interpretable features for pathologists providing effective biological signatures for histopathology diagnosis [92].

2. **Fully-integrated computerized solutions for pathology workflow.** There are lots of works proposing computerized solutions for different digital pathology tasks. However, there are very few works suggesting general frameworks or proposing integrable solutions for each state of the pathology workflow analysis, starting from tumor detection, tumor characterization and grading, diagnosis and survival prognosis.
3. **Combining learned and handcrafted features obtains better results.** The main trend in histopathology image analysis have been specialized hand-crafted features to represent different visual patterns in histopathology images. Such as it was proposed in this thesis and other preliminary works, there is a new trend based on representation learning techniques to address the same problems with successful and promising results. However, in the challenging tasks of mitosis detection in breast cancer (Section 4.1) and medulloblastoma tumor differentiation in brain cancer [131], the best performance have been achieved by combining hand-crafted features and learned features. This suggests that hybrid approaches could be a better alternative, hence more research in this direction is required for taking advantage of prior domain knowledge from hand-crafted features and implicit and hidden complex relationships of visual patterns captured by the representation learning approaches.
4. **Towards translational and personalized medicine.** This work showed that it is possible to develop effective and efficient methods for supporting pathology diagnosis and grading. However, the translation of these methods to real scenarios in clinical practice, require further work to guarantee: high-throughput analysis of whole-slide images in feasible computing time, reproducibility and the integration of different evidence sources.

# Bibliography

- [1] AMERICAN CANCER SOCIETY: Cancer facts & figures 2008. 2008. – Technical report
- [2] ALAM, Murad ; RATNER, Désirée: Cutaneous Squamous-Cell Carcinoma. In: *New England Journal of Medicine* 344 (2001), Nr. 13, 975-983. <http://dx.doi.org/10.1056/NEJM200103293441306>. – DOI 10.1056/NEJM200103293441306. – PMID: 11274625
- [3] ANGELOVA, Donka ; MIHAYLOVA, Lyudmila: Contour segmentation in 2D ultrasound medical images with particle filtering. In: *Machine Vision and Applications* 22 (2011), Nr. 3, 551-561. <http://dx.doi.org/10.1007/s00138-010-0261-4>. – DOI 10.1007/s00138-010-0261-4. – ISSN 0932-8092
- [4] AREVALO, John ; CRUZ-ROA, Angel ; ARIAS, Viviana ; ROMERO, Eduardo ; GONZALEZ, Fabio A.: An unsupervised feature learning framework for basal cell carcinoma image analysis. In: *Artificial Intelligence in Medicine* 64 (2015), Nr. 2, 131 – 145. <http://dx.doi.org/http://dx.doi.org/10.1016/j.artmed.2015.04.004>. – DOI <http://dx.doi.org/10.1016/j.artmed.2015.04.004>. – ISSN 0933-3657
- [5] AREVALO, John ; CRUZ-ROA, Angel ; GONZÁLEZ, Fabio: Hybrid image representation learning model with invariant features for basal cell carcinoma detection. In: *The 9th International Seminar on Medical Information Processing and Analysis*, 2013
- [6] ARÉVALO, John ; CRUZ-ROA, Angel ; GONZÁLEZ, Fabio: Histopathology image representation for automatic analysis: A state-of-the-art review. In: *Revista Med* 22 (2014), Nr. 2, S. 79–91. – ISSN: 0121-5256
- [7] ATUPELAGE, Chamidu ; NAGAHASHI, Hiroshi ; YAMAGUCHI, Masahiro ; ABE, Tokiya ; HASHIGUCHI, Akinori ; SAKAMOTO, Michiie: Computational grading of hepatocellular carcinoma using multifractal feature description. In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical*

- Imaging Society* 37 (2013), Januar, Nr. 1, 61–71. <http://dx.doi.org/10.1016/j.compmedimag.2012.10.001>. – ISSN 1879–0771
- [8] BAARDWIJK, Angela van ; BOSMANS, Geert ; BOERSMA, Liesbeth ; BUIJSEN, Jeroen ; WANDERS, Stofferinus ; HOCHSTENBAG, Monique ; SUYLEN, Robert-Jan van ; DEKKER, André Kadjacsy-Balla ; DEHING-OBERIJE, Cary ; HOUBEN, Ruud ; BENTZEN, Søren M. ; KROONENBURGH, Marinus van ; LAMBIN, Philippe ; RUYSSCHER, Dirk D.: PET-CT-Based Auto-Contouring in Non-Small-Cell Lung Cancer Correlates With Pathology and Reduces Interobserver Variability in the Delineation of the Primary Tumor and Involved Nodal Volumes. In: *International Journal of Radiation Oncology\*Biology\*Physics* 68 (2007), Nr. 3, 771 - 778. <http://dx.doi.org/http://dx.doi.org/10.1016/j.ijrobp.2006.12.067>. – DOI <http://dx.doi.org/10.1016/j.ijrobp.2006.12.067>. – ISSN 0360–3016
- [9] BASAVANHALLY, A ; FELDMAN, Michael D. ; SHIH, Natalie ; MIES, C ; TOMASZEWSKI, John E. ; 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: *Journal of Pathology Informatics* 2 (2011), 01/2012, Nr. 1. <http://dx.doi.org/10.4103/2153-3539.92027>. – DOI 10.4103/2153–3539.92027
- [10] BASAVANHALLY, A ; GANESAN, Shridar ; FELDMAN, Michael D. ; SHIH, Natalie ; MIES, C ; TOMASZEWSKI, John E. ; MADABHUSHI, Anant: Multi-Field-of-View Framework for Distinguishing Tumor Grade in ER+ Breast Cancer From Entire Histopathology Slides. In: *IEEE transactions on biomedical engineering* 60 (2013), Aug, Nr. 8, S. 2089–2099
- [11] BASAVANHALLY, A ; YU, E ; XU, Jun ; GANESAN, Shridar ; FELDMAN, Michael D. ; TOMASZEWSKI, John E. ; MADABHUSHI, Anant: Incorporating domain knowledge for tubule detection in breast histopathology using O’Callaghan neighborhoods. In: *SPIE Medical Imaging* Bd. 7963 SPIE, SPIE, 2011, 796310
- [12] BASAVANHALLY, Ajay N. ; GANESAN, Shridar ; AGNER, Shannon ; MONACO, James P. ; FELDMAN, Michael D. ; TOMASZEWSKI, John E. ; BHANOT, Gyan ; MADABHUSHI, Anant: Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. In: *IEEE transactions on bio-medical engineering* 57 (2010), März, Nr. 3, 642–53. <http://dx.doi.org/10.1109/TBME.2009.2035305>. – DOI 10.1109/TBME.2009.2035305. – ISSN 1558–2531



- [13] BASKAR, Rajamanickam ; LEE, Kuo A. ; YEO, Richard ; YEOH, Kheng-Wei: Cancer and Radiation Therapy: Current Advances and Future Directions. In: *International Journal of Medical Sciences* 9 (2011), Dezember, Nr. 3, 193–199. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3298009/>. – ISSN 1449–1907
- [14] BECK, Andrew H. ; SANGOI, Ankur R. ; LEUNG, Samuel ; MARINELLI, Robert J. ; NIELSEN, Torsten O. ; VIJVER, Marc J. d. ; WEST, Robert B. ; RIJN, Matt van d. ; KOLLER, Daphne: Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival. In: *Science Translational Medicine* 3 (2011), Nr. 108, 108ra113. <http://dx.doi.org/10.1126/scitranslmed.3002564>. – DOI 10.1126/scitranslmed.3002564
- [15] BENGIO, Yoshua: Learning Deep Architectures for AI. In: *Found. Trends Mach. Learn.* 2 (2009), Januar, Nr. 1, 1–127. <http://dx.doi.org/10.1561/22000000006>. – DOI 10.1561/22000000006. – ISSN 1935–8237
- [16] BENGIO, Yoshua ; COURVILLE, Aaron ; VINCENT, Pascal: Representation Learning: A Review and New Perspectives. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013), Nr. 8, S. 1798–1828. <http://dx.doi.org/http://doi.ieeecomputersociety.org/10.1109/TPAMI.2013.50>. – DOI <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2013.50>. – ISSN 0162–8828
- [17] BIEDERMAN, Irving: Recognition-by-components: A theory of human image understanding. In: *Psychological Review* 94 (1987), S. 115–147
- [18] BLEI, David M. ; NG, Andrew Y. ; JORDAN, Michael I.: Latent dirichlet allocation. In: *J. Mach. Learn. Res.* 3 (2003), March, S. 993–1022. – ISSN 1532–4435
- [19] BLOOM, H. J. ; RICHARDSON, W. W.: Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. In: *British journal of cancer* 11 (1957), Nr. 3, S. 359–77
- [20] BOTTOU, Léon ; BOUSQUET, Olivier: The Tradeoffs of Large Scale Learning. Version: 2008. <http://leon.bottou.org/papers/bottou-bousquet-2008>. In: PLATT, J.C. (Ed.) ; KOLLER, D. (Ed.) ; SINGER, Y. (Ed.) ; ROWEIS, S. (Ed.): *Advances in Neural Information Processing Systems* Bd. 20. NIPS Foundation (<http://books.nips.cc>), 2008, 161–168
- [21] BOUCHERON, Laura E.: *Object- and Spatial-Level Quantitative Analysis of Multispectral Histopathology Images for Detection and Characterization of Cancer*, Diss., Mar 2008

- [22] CAFLISCH, Russel E.: Monte carlo and quasi-monte carlo methods. In: *Acta numerica* 7 (1998), S. 1–49
- [23] CHANG, H. ; LOSS, L. ; PARVIN, B.: Nuclear segmentation in H&E sections via multi-reference graph cut (MRGC). In: *ISBI'2012: Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging*, 2012
- [24] CHANG, Hang ; HAN, Ju ; BOROWSKY, A. ; LOSS, L. ; GRAY, J.W. ; SPELLMAN, P.T. ; PARVIN, B.: Invariant Delineation of Nuclear Architecture in Glioblastoma Multiforme for Clinical and Molecular Association. In: *Medical Imaging, IEEE Transactions on* 32 (2013), April, Nr. 4, S. 670–682. <http://dx.doi.org/10.1109/TMI.2012.2231420>. – DOI 10.1109/TMI.2012.2231420. – ISSN 0278–0062
- [25] CHAPPELOW, Jonathan ; TOMASZEWSKI, John E. ; FELDMAN, Michael ; SHIH, Natalie ; MADABHUSHI, Anant: HistoStitcher(c): an interactive program for accurate and rapid reconstruction of digitized whole histological sections from tissue fragments. In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 35 (2011), Januar, Nr. 7-8, 557–67. <http://www.sciencedirect.com/science/article/pii/S0895611111000218>. – ISSN 1879–0771
- [26] CHAWLA, N. V. ; BOWYER, K. W. ; HALL, L. O. ; KEGELMEYER, W. P.: SMOTE: synthetic minority over-sampling technique. In: *Journal of Artificial Intelligence Research* 16 (2002), Nr. 1, S. 321–357
- [27] CHEN, Wei ; KOLDITZ, Daniel ; BEISTER, Marcel ; BOHLE, Robert ; KALENDER, Willi A.: Fast on-site Monte Carlo tool for dose calculations in CT applications. In: *Medical Physics* 39 (2012), Nr. 6, 2985–2996. <http://dx.doi.org/http://dx.doi.org/10.1118/1.4711748>. – DOI <http://dx.doi.org/10.1118/1.4711748>
- [28] CIRESAN, Dan. ; GIUSTI, Alessandro ; GAMBARDELLA, LucaM. ; SCHMIDHUBER, Jürgen: Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013* Bd. 8150, Springer Berlin Heidelberg, 2013 (Lecture Notes in Computer Science). – ISBN 978–3–642–40762–8, S. 411–418
- [29] COLLOBERT, R. ; KAVUKCUOGLU, K. ; FARABET, C.: Torch7: A Matlab-like Environment for Machine Learning. In: *BigLearn, NIPS Workshop*, 2011
- [30] CRUZ-ROA, A. ; DIAZ, G. ; GONZALEZ, F.: A framework for semantic analysis of histopathological images using nonnegative matrix factorization. In: *Computing Congress (CCC), 2011 6th Colombian*, 2011, S. 1 –7

- [31] CRUZ-ROA, Angel ; ARÉVALO, John ; BASAVANHALLY, Ajay ; MADABHUSHI, Anant ; GONZALEZ, Fabio: A comparative evaluation of supervised and unsupervised representation learning approaches for anaplastic medulloblastoma differentiation. In: *Tenth International Symposium on Medical Information Processing and Analysis (SIPAIM 2014)*, 2014
- [32] CRUZ-ROA, Angel ; ARÉVALO, John ; BASAVANHALLY, Ajay ; MADABHUSHI, Anant ; GONZÁLEZ, Fabio: A comparative evaluation of supervised and unsupervised representation learning approaches for anaplastic medulloblastoma differentiation, 2015, 92870G-92870G-6
- [33] CRUZ-ROA, Angel ; AREVALO, John ; MADABHUSHI, Anant ; GONZÁLEZ, Fabio: A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection. Version:2013. [http://dx.doi.org/10.1007/978-3-642-40763-5\\_50](http://dx.doi.org/10.1007/978-3-642-40763-5_50). In: MORI, Kensaku (Ed.) ; SAKUMA, Ichiro (Ed.) ; SATO, Yoshinobu (Ed.) ; BARILLOT, Christian (Ed.) ; NAVAB, Nassir (Ed.): *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013* Bd. 8150. Springer Berlin Heidelberg, 2013. – DOI 10.1007/978-3-642-40763-5\_50. – ISBN 978-3-642-40762-8, 403–410
- [34] CRUZ-ROA, Angel ; BASAVANHALLY, Ajay ; FELDMAN, Michael ; GANESAN, Shridar ; SHIH, Natalie ; TOMASZEWSKI, John ; GILMORE, Hannah ; MADABHUSHI, Anant ; GONZÁLEZ, Fabio: Scaling up convolutional networks through high-throughput adaptive sampling for whole-slide histopathology images (HASHI): Application to invasive breast cancer tumor detection. In: *IEEE Transactions on Medical Imaging* (2015). – ISSN 0278-0062. – (under review)
- [35] CRUZ-ROA, Angel ; BASAVANHALLY, Ajay ; GONZALEZ, Fabio ; FELDMAN, Michael ; GANESAN, Shridar ; SHIH, Natalie ; TOMASZEWSKI, John ; GILMORE, Hannah ; ; MADABHUSHI, Anant: A Feature Learning Framework for Reproducible Invasive Tumor Detection of Breast Cancer in Whole-Slide Images. In: *USCAP 104th Annual Meeting, In Labortory Investigation* Bd. 95. New York. USA : Nature Publishing Group, March 2015, S. 40A–40A
- [36] CRUZ-ROA, Angel ; BASAVANHALLY, Ajay ; GONZALEZ, 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, 2014, 904103-904103-15
- [37] CRUZ-ROA, Angel ; CAICEDO, Juan C. ; GONZÁLEZ, Fabio: Visual Mining in Histology Images Using Bag of Features. In: *6th International Seminar on*

- Medical Image Processing and Analysis - SIPAIM 2010*. Hemeroteca Nacional de Colombia. Bogotá (Colombia), December 2010
- [38] CRUZ-ROA, Angel ; CAICEDO, Juan C. ; GONZÁLEZ, Fabio: Visual Pattern Mining in Histology Image Collections Using Bag of Features. In: *Journal Artificial Intelligence in Medicine* 52 (2011), June, Nr. 2, S. 91–106. <http://dx.doi.org/10.1016/j.artmed.2011.04.010>. – DOI 10.1016/j.artmed.2011.04.010
- [39] CRUZ-ROA, Angel ; DÍAZ, Gloria ; ROMERO, Eduardo ; FABIO, González: Automatic Annotation of Histopathological Images Using a Latent Topic Model Based On Non-negative Matrix Factorization. In: *MICCAI 2011 Workshop Proceedings*. Canada, September 2011
- [40] CRUZ-ROA, Angel ; DÍAZ, Gloria ; ROMERO, Eduardo ; GONZÁLEZ, Fabio: Automatic annotation of histopathological images using a latent topic model based on non-negative matrix factorization. In: *Journal of pathology informatics* 2 (2011), Januar, Nr. 1, S4. <http://dx.doi.org/10.4103/2153-3539.92031>. – DOI 10.4103/2153-3539.92031. – ISSN 2153-3539
- [41] CRUZ-ROA, Angel ; GILMORE, Hannah ; BASAVANHALLY, Ajay ; FELDMAN, Michael ; GANESAN, Shridar ; SHIH, Natalie ; TOMASZEWSKI, John ; GONZÁLEZ, Fabio ; MADABHUSHI, Anant: Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning based tool for quantifying tumor location and extent. In: *Journal of the American Medical Informatics Association* (2015). – ISSN 0893-3952. – (under review)
- [42] CRUZ-ROA, Angel ; GONZÁLEZ, Fabio ; GALARO, Joseph ; JUDKINS, Alexander R. ; ELLISON, David ; BACCON, Jennifer ; MADABHUSHI, Anant ; ROMERO, Eduardo: A Visual Latent Semantic Approach for Automatic Analysis and Interpretation of Anaplastic Medulloblastoma Virtual Slides. In: *15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'2012)* (2012), S. 157–164
- [43] CSURKA, Gabriella ; DANCE, Christopher R. ; FAN, Lixin ; WILLAMOWSKI, Jutta ; BRAY, Cédric: Visual categorization with bags of keypoints. In: *ECCV International Workshop on Statistical Learning in Computer Vision*. Prague, 2004, S. 1–22
- [44] 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
- [45] DANIEL, Christel ; ROJO, Marcial G. ; KLOSSA, Jacques ; DELLA MEA, Vincenzo ; BOOKER, David ; BECKWITH, Bruce A. ; SCHRADER, Thomas: Standardizing the use of whole slide images in digital pathology. In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 35 (2011), Januar, Nr. 7-8, 496–505. <http://www.sciencedirect.com/science/article/pii/S08956111110001357>. – ISSN 1879–0771
- [46] DENG, Jia ; DONG, Wei ; SOCHER, R. ; LI, Li-Jia ; LI, Kai ; FEI-FEI, Li: ImageNet: A large-scale hierarchical image database. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009. – ISSN 1063–6919, S. 248–255
- [47] DENG, Li ; YU, Dong: Deep Learning: Methods and Applications. In: *Foundations and Trends in Signal Processing* 7 (2014), Nr. 3-4, 197–387. <http://dx.doi.org/10.1561/20000000039>. – DOI 10.1561/20000000039. – ISSN 1932–8346
- [48] DESANTIS, Carol ; SIEGEL, Rebecca ; BANDI, Priti ; JEMAL, Ahmedin: Breast cancer statistics, 2011. In: *CA: A Cancer Journal for Clinicians* 61 (2011), Nr. 6, 408–418. <http://dx.doi.org/10.3322/caac.20134>. – DOI 10.3322/caac.20134. – ISSN 1542–4863
- [49] DESELAERS, Thomas ; FERRARI, Vittorio: Global and Efficient Self-Similarity for Object Classification and Detection. In: *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*. Los Alamitos, CA, USA : IEEE Computer Society, June 2010 (CVPR 2010). – ISBN 978–1–4244–6984–0, S. 1633–1640
- [50] DÍAZ, Gloria ; ROMERO, Eduardo: Histopathological Image Classification Using Stain Component Features on a pLSA Model. In: BLOCH, Isabelle (Ed.) ; CESAR, Roberto (Ed.): *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* Bd. 6419. Springer Berlin / Heidelberg, 2010, S. 55–62
- [51] DÍAZ, Gloria ; ROMERO, Eduardo: Micro-structural tissue analysis for automatic histopathological image annotation. In: *Microscopy research and technique* 75 (2011), März, Nr. 3, 343–358. <http://dx.doi.org/10.1002/jemt.21063>. – DOI 10.1002/jemt.21063. – ISSN 1097–0029

- [52] DIEPGEN, TL ; MAHLER, V: The epidemiology of skin cancer. In: *British Journal of Dermatology* 146 (2002), Nr. s61, S. 1–6
- [53] DILLON DA, Schnitt S. Guidi AJ A. Guidi AJ: Pathology of Invasive Breast Cancer. In: HARRIS JR, Morrow M Osborne C. Lippman ME M. Lippman ME (Ed.): *Diseases of the Breast*. 4th edition. Lippincott Williams & Wilkins, 2010, Kapitel Chapter 28, S. 374–407
- [54] DING, Chris ; LI, Tao ; JORDAN, Michael I.: Convex and Semi-Nonnegative Matrix Factorizations. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (2010), January, 45–55. <http://dx.doi.org/http://dx.doi.org/10.1109/TPAMI.2008.277>. – DOI <http://dx.doi.org/10.1109/TPAMI.2008.277>. – ISSN 0162–8828
- [55] DING, Chris ; LI, Tao ; PENG, Wei: On the equivalence between Non-negative Matrix Factorization and Probabilistic Latent Semantic Indexing. In: *Computational Statistics & Data Analysis* 52 (2008), April, Nr. 8, 3913–3927. <http://dx.doi.org/10.1016/j.csda.2008.01.011>. – DOI 10.1016/j.csda.2008.01.011. – ISSN 01679473
- [56] DONAHUE, Jeff ; JIA, Yangqing ; VINYALS, Oriol ; HOFFMAN, Judy ; ZHANG, Ning ; TZENG, Eric ; DARRELL, Trevor: Decaf: A deep convolutional activation feature for generic visual recognition. In: *arXiv preprint arXiv:1310.1531* (2013)
- [57] DORAI, C. ; JAIN, A.K.: COSMOS-A representation scheme for 3D free-form objects. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19 (1997), Oct, Nr. 10, S. 1115–1130. <http://dx.doi.org/10.1109/34.625113>. – DOI 10.1109/34.625113. – ISSN 0162–8828
- [58] DOYLE, Scott ; AGNER, Shannon ; MADABHUSHI, Anant ; FELDMAN, Michael ; TOMASZEWSKI, John: Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In: *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, Mai 2008. – ISBN 978–1–4244–2002–5, 496–499
- [59] DOYLE, Scott ; MADABHUSHI, Anant ; FELDMAN, Michael ; TOMASZEWSKI, John: A Boosting Cascade for Automated Detection of Prostate Cancer from Digitized Histology. In: LARSEN, Rasmus (Ed.) ; NIELSEN, Mads (Ed.) ; SPORRING, Jon (Ed.): *Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part II*. Berlin, Heidelberg : Springer-Verlag, 2006 (MICCAI’06). – ISBN 3–540–44727–X, 978–3–540–44727–6, S. 504–511

- [60] DUNDAR, M. M. ; BADVE, Sunil ; BILGIN, Gokhan ; RAYKAR, Vikas ; JAIN, Rohit ; SERTEL, Olcay ; GURCAN, Metin N.: Computerized classification of intraductal breast lesions using histopathological images. In: *IEEE Transactions on Biomedical Engineering* 58 (2011), Juli, Nr. 7, 1977–1984. <http://dx.doi.org/10.1109/TBME.2011.2110648>. – DOI 10.1109/TBME.2011.2110648. – ISSN 00189294
- [61] ELMORE, Joann G. ; LONGTON, Gary M. ; CARNEY, Patricia a. ; GELLER, Berta M. ; ONEGA, Tracy ; TOSTESON, Anna N a. ; NELSON, Heidi D. ; PEPE, Margaret S. ; ALLISON, Kimberly H. ; SCHNITT, Stuart J. ; MALLEY, Frances P O. ; WEAVER, Donald L.: Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. 98104 (2015), Nr. 11, S. 1122–1132. <http://dx.doi.org/10.1001/jama.2015.1405>. – DOI 10.1001/jama.2015.1405
- [62] 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), S. 403–10
- [63] ESSERMAN, LJ ; THOMPSON, IM ; REID, B: Overdiagnosis and overtreatment in cancer: An opportunity for improvement. In: *JAMA* 310 (2013), Nr. 8, 797–798. <http://dx.doi.org/10.1001/jama.2013.108415>. – DOI 10.1001/jama.2013.108415
- [64] FERLAY, Jacques ; SOERJOMATARAM, Isabelle ; DIKSHIT, Rajesh ; ESER, Sultan ; MATHERS, Colin ; REBELO, Marise ; PARKIN, Donald M. ; FORMAN, David ; BRAY, Freddie: Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. In: *International Journal of Cancer* 136 (2015), Nr. 5, E359–E386. <http://dx.doi.org/10.1002/ijc.29210>. – DOI 10.1002/ijc.29210. – ISSN 1097–0215
- [65] FLETCHER, Christopher D. M. ; FLETCHER, Christopher D. M. (Ed.): *Diagnostic Histopathology of tumors*. Elsevier Science, 2003
- [66] FLORIN, Charles ; PARAGIOS, Nikos ; WILLIAMS, Jim: Particle Filters, a Quasi-Monte Carlo Solution for Segmentation of Coronaries. Version:2005. [http://dx.doi.org/10.1007/11566465\\_31](http://dx.doi.org/10.1007/11566465_31). In: DUNCAN, JamesS. (Ed.) ; GERIG, Guido (Ed.): *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2005* Bd. 3749. Springer Berlin Heidelberg, 2005. – ISBN 978–3–540–29327–9, 246–253

- [67] FRIERSON, H F. ; WOLBER, R A. ; BEREAN, K W. ; FRANQUEMONT, D W. ; GAFFEY, M J. ; BOYD, J C. ; WILBUR, D C.: Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. In: *American journal of clinical pathology* 103 (1995), Februar, Nr. 2, 195–8. <http://europepmc.org/abstract/med/7856562>. – ISSN 0002–9173
- [68] FUCHS, Thomas J. ; BUHMANN, Joachim M.: Computational pathology: Challenges and promises for tissue analysis. In: *Comp. Med. Imag. and Graph.* 35 (2011), Nr. 7-8, S. 515–530
- [69] GALARO, J. ; JUDKINS, A.R. ; ELLISON, D. ; BACCON, J. ; MADABHUSHI, A.: An integrated textron and bag of words classifier for identifying anaplastic medulloblastomas. In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011. – ISSN 1557–170X, S. 3443–3446
- [70] GALLOWAY, M. M.: Texture analysis using gray level run lengths. In: *Computer Graphics and Image Processing* 4 (1975), Nr. 2, S. 172 – 179
- [71] GENESTIE, C. ; ZAFRANI, B. ; ASSELAIN, B. ; FOURQUET, A. ; ROZAN, S. ; VALIDIRE, P. ; VINCENT-SALOMON, A. ; SASTRE-GARAU, X.: Comparison of the prognostic value of Scarff-Bloom-Richardson and nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems. In: *Anticancer Research* 18 (1998), Nr. 1B, S. 571–576
- [72] GOMES, Douglas S. ; PORTO, Simone S. ; BALABRAM, Débora ; GOBBI, Helenice: Inter-observer variability between general pathologists and a specialist in breast pathology in the diagnosis of lobular neoplasia, columnar cell lesions, atypical ductal hyperplasia and ductal carcinoma in situ of the breast. In: *Diagnostic pathology* 9 (2014), Januar, 121. <http://dx.doi.org/10.1186/1746-1596-9-121>. – DOI 10.1186/1746–1596–9–121. – ISSN 1746–1596
- [73] GURCAN, Metin N. ; BOUCHERON, Laura E. ; CAN, Ali ; MADABHUSHI, Anant ; RAJPOOT, Nasir M. ; YENER, Bulent: Histopathological image analysis: A review. In: *Biomedical Engineering, IEEE Reviews in* 2 (2009), S. 147–171
- [74] GUTMAN, David A. ; COBB, Jake ; SOMANNA, Dhananjaya ; PARK, Yuna ; WANG, Fusheng ; KURC, Tahsin ; SALTZ, Joel H. ; BRAT, Daniel J. ; COOPER, Lee A D. ; KONG, Jun: Cancer Digital Slide Archive: an informatics resource



- to support integrated in silico analysis of TCGA pathology data. In: *Journal of the American Medical Informatics Association* 20 (2013), Nr. 6, S. 1091–1098. <http://dx.doi.org/10.1136/amiajnl-2012-001469>. – DOI 10.1136/amiajnl-2012-001469. – ISBN 1527–974X
- [75] HARALICK, R. M. ; SHAPIRO, L. G.: *Computer and robot vision*. Boston, MA, USA : Addison-Wesley Publishing Company, 1992
- [76] HARALICK, Robert M. ; SHANMUGAM, K. ; DINSTEN, Its'Hak: Textural Features for Image Classification. In: *IEEE Transactions on Systems, Man, and Cybernetics* 3 (1973), November, Nr. 6, 610–621. <http://dx.doi.org/10.1109/TSMC.1973.4309314>. – DOI 10.1109/TSMC.1973.4309314. – ISSN 0018–9472
- [77] HINTON, GE ; SRIVASTAVA, N: Improving neural networks by preventing co-adaptation of feature detectors. In: *arXiv preprint arXiv: ...* (2012). <http://arxiv.org/abs/1207.0580>
- [78] HIPPEL, Jason D. ; SMITH, Steven C. ; SICA, Jeffrey. ; LUCAS, David. ; HIPPEL, Jennifer A. ; KUNJU, Lakshmi P. ; BALIS, Ulysses J.: Tryggo: Old nose for truth: The real truth about ground truth: New insights into the challenges of generating ground truth maps for WSI CAD algorithm evaluation. In: *Journal of pathology informatics* 3 (2012), Januar, Nr. 1, 8. <http://dx.doi.org/10.4103/2153-3539.93890>. – DOI 10.4103/2153-3539.93890. – ISSN 2153–3539
- [79] HITCHCOCK, Charles L.: The Future of Telepathology for the Developing World. In: *Archives of Pathology & Laboratory Medicine* 135 (2011), Nr. 2, 211–214. <http://dx.doi.org/10.1043/1543-2165-135.2.211>. – DOI 10.1043/1543-2165-135.2.211. – PMID: 21284440
- [80] HOFMANN, Thomas: Unsupervised Learning by Probabilistic Latent Semantic Analysis. In: *Mach. Learn.* 42 (2001), January, S. 177–196. <http://dx.doi.org/10.1023/A:1007617005950>. – DOI 10.1023/A:1007617005950. – ISSN 0885–6125
- [81] HOMEYER, André ; SCHENK, Andrea ; ARLT, Janine ; DAHMEN, Uta ; DIRSCH, Olaf ; HAHN, Horst K.: Practical quantification of necrosis in histological whole-slide images. In: *Computerized Medical Imaging and Graphics* (2013), Juni, 313–322. <http://dx.doi.org/10.1016/j.compmedimag.2013.05.002>. – DOI 10.1016/j.compmedimag.2013.05.002. – ISSN 08956111

- [82] HUANG, C.-H. ; LEE, H.-K.: Automated mitosis detection based on eXclusive Independent Component Analysis. In: *International Conference on Pattern Recognition (ICPR)*, 2012, S. 1856–1859
- [83] HUANG, Chao-Hui ; VEILLARD, Antoine ; ROUX, Ludovic ; LOMÉNIE, Nicolas ; RACOCEANU, Daniel: Time-efficient sparse analysis of histopathological whole slide images. In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 35 (2011), Januar, Nr. 7-8, 579–91. <http://dx.doi.org/10.1016/j.compmedimag.2010.11.009>. – DOI 10.1016/j.compmedimag.2010.11.009. – ISSN 1879–0771
- [84] HYVÄRINEN, Aapo ; HURRI, Jarmo ; HOYER, Patrik O.: *Natural image statistics*. Bd. 39. Springer, 2009
- [85] IGLESIAS, Juan E. ; SABUNCU, Mert R. ; VAN LEEMPUT, Koen: Improved inference in Bayesian segmentation using Monte Carlo sampling: Application to hippocampal subfield volumetry. In: *Medical Image Analysis* 17 (2013), März, Nr. 7, 766–778. <http://dx.doi.org/10.1016/j.media.2013.04.005>. – DOI 10.1016/j.media.2013.04.005
- [86] IRSHAD, H ; JALALI, S ; ROUX, L ; RACOCEANU, D. ; HWEE, L J. ; NAOUR, G L. ; CAPRON, F: Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach. In: *Journal of Pathology Informatics* 4 (2013), Nr. 2, S. 12
- [87] IRSHAD, Humayun: Automated Mitosis Detection in Histopathology using Morphological and Multi-channels Statistics Features. In: *Journal of Pathology Informatics* 4 (2013), 05/2013. <http://dx.doi.org/10.4103/2153-3539.112695>. – DOI 10.4103/2153–3539.112695
- [88] JIA, Xun ; YAN, Hao ; GU, Xuejun ; JIANG, Steve B.: Fast Monte Carlo simulation for patient-specific CT/CBCT imaging dose calculation. In: *Physics in Medicine and Biology* 57 (2012), Nr. 3, 577. <http://stacks.iop.org/0031-9155/57/i=3/a=577>
- [89] KAMIYA, Yasunori ; TAKAHASHI, Tomokazu ; IDE, Ichiro ; MURASE, Hiroshi: A Multimodal Constellation Model for Object Category Recognition. In: HUET, Benoit (Ed.) ; SMEATON, Alan (Ed.) ; MAYER-PATEL, Ketan (Ed.) ; AVRITHIS, Yannis (Ed.): *Advances in Multimedia Modeling* Bd. 5371. Springer Berlin / Heidelberg, 2009, S. 310–321

- [90] KOTHARI, S. ; PHAN, J.H. ; STOKES, T.H. ; OSUNKOYA, A.O. ; YOUNG, A.N. ; WANG, M.D.: Removing Batch Effects From Histopathological Images for Enhanced Cancer Diagnosis. In: *Biomedical and Health Informatics, IEEE Journal of* 18 (2014), May, Nr. 3, S. 765–772. <http://dx.doi.org/10.1109/JBHI.2013.2276766>. – DOI 10.1109/JBHI.2013.2276766. – ISSN 2168–2194
- [91] KOTHARI, Sonal ; OSUNKOYA, Adeboye O. ; PHAN, John H. ; WANG, May D.: Biological interpretation of morphological patterns in histopathological whole-slide images. In: *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine - BCB '12*. New York, New York, USA : ACM Press, 2012. – ISBN 9781450316705, 218–225
- [92] KOTHARI, Sonal ; PHAN, John H. ; STOKES, Todd H. ; WANG, May D.: Pathology imaging informatics for quantitative analysis of whole-slide images. In: *Journal of the American Medical Informatics Association* 20 (2013), Nr. 6, 1099–1108. <http://dx.doi.org/10.1136/amiajnl-2012-001540>. – DOI 10.1136/amiajnl-2012-001540
- [93] KRIZHEVSKY, Alex ; SUTSKEVER, Ilya ; HINTON, Geoffrey E.: ImageNet Classification with Deep Convolutional Neural Networks. Version: 2012. [http://books.nips.cc/papers/files/nips25/NIPS2012\\_0534.pdf](http://books.nips.cc/papers/files/nips25/NIPS2012_0534.pdf). In: *Advances in Neural Information Processing Systems 25*. 2012, 1106–1114
- [94] KVILEKVAL, Kristian ; FEDOROV, Dmitry ; OBARA, Boguslaw ; SINGH, Ambuj ; MANJUNATH, B. S.: Bisque: a platform for bioimage analysis and management. In: *Bioinformatics* 26 (2010), Februar, Nr. 4, S. 544–552. <http://dx.doi.org/10.1093/bioinformatics/btp699>. – DOI 10.1093/bioinformatics/btp699
- [95] KWAK, Jin T. ; HEWITT, Stephen M. ; SINHA, Saurabh ; BHARGAVA, Rohit: Multimodal microscopy for automated histologic analysis of prostate cancer. In: *BMC cancer* 11 (2011), Nr. 1, 62. <http://dx.doi.org/10.1186/1471-2407-11-62>. – DOI 10.1186/1471-2407-11-62. – ISSN 1471–2407
- [96] LACROIX-TRIKI, Magali ; SUAREZ, Paula H. ; MACKAY, Alan ; LAMBROS, Maryou B. ; NATRAJAN, Rachael ; SAVAGE, Kay ; GEYER, Felipe C. ; WEIGELT, Britta ; ASHWORTH, Alan ; REIS-FILHO, Jorge S.: Mucinous carcinoma of the breast is genomically distinct from invasive ductal carcinomas of no special type. In: *The Journal of Pathology* 222 (2010), Nr. 3, 282–298. <http://dx.doi.org/10.1002/path.2763>. – DOI 10.1002/path.2763. – ISSN 1096–9896

- [97] LE, Quoc ; NGIAM, Jiquan ; CHEN, Zhenghao ; CHIA, Daniel J. ; KOH, Pang W. ; NG, Andrew: Tiled convolutional neural networks. In: *Advances in Neural Information Processing Systems 23*. 2010, S. 1279–1287
- [98] LE, Quoc ; RANZATO, Marc'Aurelio ; MONGA, Rajat ; DEVIN, Matthieu ; CHEN, Kai ; CORRADO, Greg ; DEAN, Jeff ; NG, Andrew: Building high-level features using large scale unsupervised learning. In: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (2011)
- [99] LE, Quoc ; RANZATO, Marc'Aurelio ; MONGA, Rajat ; DEVIN, Matthieu ; CHEN, Kai ; CORRADO, Greg ; DEAN, Jeff ; NG, Andrew: Building high-level features using large scale unsupervised learning. In: *International Conference in Machine Learning*, 2012
- [100] LE, Quoc V. ; HAN, Ju ; GRAY, Joe W. ; SPELLMAN, Paul T. ; BOROWSKY, Alexander ; PARVIN, Bahram: Learning invariant features of tumor signatures. In: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on* IEEE, 2012, S. 302–305
- [101] LECUN, Yann: Learning invariant feature hierarchies. In: FUSIELLO, Andrea (Ed.) ; MURINO, Vittorio (Ed.) ; CUCCHIARA, Rita (Ed.) ; Springer (Veranst.): *Computer Vision–ECCV 2012. Workshops and Demonstrations* Springer, Springer Berlin Heidelberg, 2012, 496–505
- [102] LECUN, Yann ; BOTTOU, Léon ; BENGIO, Yoshua ; HAFFNER, Patrick: Gradient-based learning applied to document recognition. In: *Proceedings of the IEEE* 86 (1998), Nr. 11, 2278–2324. <http://dx.doi.org/10.1109/5.726791>. – DOI 10.1109/5.726791. – ISSN 00189219
- [103] LEE, Daniel D. ; SEUNG, H. S.: Learning the parts of objects by non-negative matrix factorization. In: *Nature* 401 (1999), Nr. 6755, 788–91. <http://dx.doi.org/10.1038/44565>. – DOI 10.1038/44565. – ISSN 00280836
- [104] LEE, Daniel D. ; SEUNG, H. S.: Algorithms for Non-negative Matrix Factorization. In: *NIPS*, 2000, 556–562
- [105] LEE, George ; SPARKS, Rachel ; ALI, Sahirzeeshan ; SHIH, Natalie N C. ; FELDMAN, Michael D. ; SPANGLER, Elaine ; REBBECK, Timothy ; TOMASZEWSKI, John E. ; MADABHUSHI, Anant: Co-occurring gland angularity in localized sub-graphs: predicting biochemical recurrence in intermediate-risk prostate cancer patients. In: *PloS one* 9 (2014), 05/2014, S. e97954. <http://dx.doi.org/10>.

- 1371/journal.pone.0097954. – DOI 10.1371/journal.pone.0097954. – ISSN 1932–6203
- [106] LEI HE, Sameer A. L. Rodney Long L. L. Rodney Long ; THOMA, George R.: Histology image analysis for carcinoma detection and grading. In: *Computer Methods and Programs in Biomedicine* 107 (2012), September, Nr. 3, S. 538–556
- [107] LEIN, Ed S. ; HAWRYLYCZ, Michael J. ; AO, Nancy ; AYRES, Mikael ; BENSINGER, Amy ; AL. et: Genome-wide atlas of gene expression in the adult mouse brain. In: *Nature* 445 (2006), Dezember, Nr. 7124, 168–176. <http://dx.doi.org/10.1038/nature05453>. – DOI 10.1038/nature05453. – ISSN 0028–0836
- [108] LEWIS, James S. ; ALI, Sahirzeeshan ; LUO, Jingqin ; THORSTAD, Wade L. ; MADABHUSHI, Anant: A Quantitative Histomorphometric Classifier ( QuHbIC ) Oropharyngeal Squamous Cell Carcinoma. In: *Am J Surg Pathol* 38 (2014), Nr. 1, S. 128–137
- [109] LI, Fei-Fei ; PERONA, Pietro: A Bayesian Hierarchical Model for Learning Natural Scene Categories. In: *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*. Washington, DC, USA : IEEE Computer Society, 2005, 524–531
- [110] LOMENIE, Nicolas ; RACOCEANU, Daniel: Spatial relationships over sparse representations. In: *2009 24th International Conference Image and Vision Computing New Zealand*, IEEE, November 2009. – ISBN 978–1–4244–4697–1, 226–230
- [111] LONGACRE, Teri A. ; ENNIS, Marguerite ; QUENNEVILLE, Louise A. ; BANE, Anita L. ; BLEIWEISS, Ira J. ; CARTER, Beverley A. ; CATELANO, Edison ; HENDRICKSON, Michael R. ; HIBSHOOSH, Hanina ; LAYFIELD, Lester J. ; MEMEO, Lorenzo ; WU, Hong ; O'MALLEY, Frances P.: Interobserver agreement and reproducibility in classification of invasive breast carcinoma: an NCI breast cancer family registry study. In: *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 19 (2006), Februar, Nr. 2, 195–207. <http://dx.doi.org/10.1038/modpathol.3800496>. – DOI 10.1038/modpathol.3800496. – ISSN 0893–3952
- [112] LUDOVIC, Roux ; DANIEL, Racocceanu ; MARIA, Kulikova ; HUMAYUN, Irshad ; JACQUES, Klossa ; FRÉDÉRIQUE, Capron ; CATHERINE, Genestie ; GILLES, LeNaour ; NICOLAS, Loménie ; METIN, NGurcan: Mitosis detection in breast

- cancer histological images An ICPR 2012 contest. In: *Journal of Pathology Informatics* 4 (2013), Nr. 1, 8. <http://dx.doi.org/10.4103/2153-3539.112693>. – DOI 10.4103/2153-3539.112693. – ISSN 2153-3539
- [113] MADABHUSHI, Anant: Digital pathology image analysis: opportunities and challenges (Editorial). In: *Imaging In Medicine* 1 (2009), October, Nr. 1, S. 7–10
- [114] MADABHUSHI, Anant ; AGNER, Shannon ; BASAVANHALLY, Ajay ; DOYLE, Scott ; LEE, George ; MADABHUSHI A., Agner S Basavanhally A Doyle S Lee G. ; MADABHUSHI A., Basavanhally A Doyle S Agner S Lee G.: Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data. In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 35 (2011), Nr. 7-8, 506–14. <http://dx.doi.org/10.1016/j.compmedimag.2011.01.008>. – DOI 10.1016/j.compmedimag.2011.01.008. – ISBN 9781424441266
- [115] MAIRAL, Julien ; BACH, Francis ; PONCE, Jean ; SAPIRO, Guillermo: *Online dictionary learning for sparse coding*. New York, New York, USA : ACM Press, 2009 (ICML '09). – 1–8 S. <http://dx.doi.org/10.1145/1553374.1553463>. <http://dx.doi.org/10.1145/1553374.1553463>. – ISBN 9781605585161
- [116] MALON, CD ; COSATTO, Eric: Classification of mitotic figures with convolutional neural networks and seeded blob features. In: *Journal of Pathology Informatics* 4 (2013), Nr. 1, S. 9
- [117] MCGEE, J.O.D. ; ISAACSON, P.G. ; WRIGHT, N.A.: *Oxford Textbook of Pathology: Principles of pathology*. Oxford University Press, 1992 (Oxford medical publications). <http://books.google.com/books?id=ZvDDQgAACAAJ>. – ISBN 9780192619723
- [118] MCLENDON, Roger ; FRIEDMAN, Allan ; BIGNER, Darrell ; VAN MEIR, Erwin G. ; BRAT, Daniel J. ; MASTROGIANAKIS, Gena M. ; OLSON, Jeffrey J. ; MIKKELSEN, Tom ; LEHMAN, Norman ; ALDAPE, Ken u.a.: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. In: *Nature* 455 (2008), Nr. 7216, S. 1061–1068
- [119] MELAMED, Jonathan ; DATTA, Milton W. ; BECICH, Michael J. ; ORENSTEIN, Jan M. ; DHIR, Rajiv ; SILVER, Sylvia ; FIDÉLIA-LAMBERT, Marie ; KADJACSYPALLA, André ; MACIAS, Virgilia ; PATEL, Ashokkumar ; WALDEN, Paul D. ; BOSLAND, Maarten C. ; BERMAN, Jules J.: The Cooperative Prostate Cancer Tissue Resource: A Specimen and Data Resource for Cancer Researchers. In:

- Clinical Cancer Research* 10 (2004), Nr. 14, 4614-4621. <http://dx.doi.org/10.1158/1078-0432.CCR-04-0240>. – DOI 10.1158/1078-0432.CCR-04-0240
- [120] MONACO, James P. ; HIPPEL, Jason ; LUCAS, D ; SMITH, Steven C. ; BALIS, Ulysses ; MADABHUSHI, Anant: Image segmentation with implicit color standardization using spatially constrained expectation maximization: detection of nuclei. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 15 (2012), 2012, S. 365–72
- [121] MONTAVON, Grégoire: *A Machine Learning Approach to Classification of Low Resolution Histological Samples*, École Polytechnique Fédérale de Lausanne (EPFL, Switzerland), Diplomarbeit, 2009
- [122] MOROKOFF, William J. ; CAFLISCH, Russel E.: Quasi-random sequences and their discrepancies. In: *SIAM Journal on Scientific Computing* 15 (1994), Nr. 6, S. 1251–1279
- [123] MUELLER, Dan ; VOSSEN, Dirk ; HULSKEN, Bas: Real-time deformable registration of multi-modal whole slides for digital pathology. In: *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 35 (2011), Januar, Nr. 7-8, 542–56. <http://www.sciencedirect.com/science/article/pii/S089561111100084X>. – ISSN 1879-0771
- [124] NAIK, Shivang ; DOYLE, Scott ; AGNER, Shannon ; MADABHUSHI, Anant ; FELDMAN, Michael ; TOMASZEWSKI, John: Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In: *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, Mai 2008. – ISBN 978-1-4244-2002-5, 284–287
- [125] NETWORK, Scottish Intercollegiate G.: Management of breast cancer in women. In: *A national clinical guideline* 84 (2005), December, S. 1–3
- [126] NETWORK, The Cancer Genome Atlas R. ; WEINSTEIN, John N. ; COLLISSE, Eric A. ; MILLS, Gordon B. ; SHAW, Kenna R M. ; OZENBERGER, Brad A. ; ELLROTT, Kyle ; SHMULEVICH, Ilya ; SANDER, Chris ; STUART, Joshua M.: The Cancer Genome Atlas Pan-Cancer analysis project. In: *Nat Genet* 45 (2013), Oktober, Nr. 10, 1113–1120. <http://dx.doi.org/10.1038/ng.2764>. – ISSN 1061-4036
- [127] NIEDERREITER, Harald ; RANDOM NUMBER GENERATION, NSF-CBMS Regional C.: *Random number generation and quasi-Monte Carlo methods*. Bd. 63. Philadelphia : Society for Industrial and Applied mathematics (SIAM), 1992

- [128] NOWAK, Eric ; JURIE, Frédéric ; TRIGGS, Bill: Sampling Strategies for Bag-of-Features Image Classification. Version: 2006. [http://dx.doi.org/http://dx.doi.org/10.1007/11744085\\_38](http://dx.doi.org/http://dx.doi.org/10.1007/11744085_38). In: LEONARDIS, Aleš (Ed.) ; BISCHOF, Horst (Ed.) ; PINZ, Axel (Ed.): *Computer Vision – ECCV 2006* Bd. 3954. Springer Berlin / Heidelberg, 2006. – DOI [http://dx.doi.org/10.1007/11744085\\_38](http://dx.doi.org/10.1007/11744085_38), S. 490–503
- [129] OHBUCHI, Ryutarou ; AONO, Masaki: Quasi-Monte Carlo rendering with adaptive sampling. In: *IBM Tokyo Research Laboratory* (1996)
- [130] *Kapitel 108.* In: OLSHAUSEN, Bruno: *Principles of Image Representation in Visual Cortex*. 2003
- [131] OTÁLORA, Sebastian ; CRUZ-ROA, Angel ; ARÉVALO, John ; ATZORI, Manfredo ; MADABHUSHI, Anant ; JUDKINS, Alexander R. ; GONZÁLEZ, Fabio ; MÜLLER, Henning ; DEPEURSINGE, Adrien: Anaplastic medulloblastoma tumor differentiation by combining unsupervised feature learning and Riesz wavelets for histopathology image representation. In: *MICCAI 2015, the 18th International Conference on Medical Image Computing and Computer Assisted Intervention*. Munich (Germany), October 2015. – (Accepted)
- [132] OZDEMIR, Erdem ; SOKMENSUER, Cenk ; GUNDUZ-DEMIR, Cigdem: Histopathological image classification with the bag of words model. In: *2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU)*, IEEE, April 2011. – ISBN 978-1-4577-0462-8, 634–637
- [133] PANG, Baochuan ; ZHANG, Yi ; CHEN, Qianqing ; GAO, Zhifan ; PENG, Qinmu ; YOU, Xinge: Cell Nucleus Segmentation in Color Histopathological Imagery Using Convolutional Networks. In: TANG, Yuanyan (Ed.) ; SUEN, Ching Y. (Ed.) ; HE, Hangen (Ed.) ; IEEE (Veranst.): *Pattern Recognition (CCPR), Chinese Conference on*. Chongqing, China : Institute of Electrical & Electronics Engineers (IEEE), 2010, S. 1–5
- [134] PEARSON, K.: On lines and planes of closest fit to systems of points in space. In: *Philosophical Magazine* 2 (1901), S. 559–572
- [135] PENG, Hanchuan: Bioimage informatics: a new area of engineering biology. In: *Bioinformatics* 24 (2008), September, Nr. 17, S. 1827–1836. – ISSN 1367-4811
- [136] PEREZ, Amanda A. ; BALABRAM, Débora ; SALLES, Marcio de A. ; GOBBI, Helenice: Consultation in breast surgical pathology: interobserver diagnostic



- variability of atypical intraductal proliferative lesions. In: *Revista brasileira de ginecologia e obstetrícia : revista da Federação Brasileira das Sociedades de Ginecologia e Obstetrícia* 35 (2013), April, Nr. 4, 164–70. <http://www.ncbi.nlm.nih.gov/pubmed/23752581>. – ISSN 1806–9339
- [137] PHAN, J.H. ; QUO, C.F. ; CHENG, C. ; WANG, M.D.: Multiscale Integration of -Omic, Imaging, and Clinical Data in Biomedical Informatics. In: *Biomedical Engineering, IEEE Reviews in* 5 (2012), S. 74–87. <http://dx.doi.org/10.1109/RBME.2012.2212427>. – DOI 10.1109/RBME.2012.2212427. – ISSN 1937–3333
- [138] RAINA, Rajat ; BATTLE, Alexis ; LEE, Honglak ; PACKER, Benjamin ; NG, Andrew Y.: Self-taught Learning: Transfer Learning from Unlabeled Data. In: *Proceedings of the 24th International Conference on Machine Learning*. New York, NY, USA : ACM, 2007 (ICML '07). – ISBN 978–1–59593–793–3, 759–766
- [139] RAZA, S ; PARRY, R ; MOFFITT, R: An Analysis of Scale and Rotation Invariance in the Bag-of-Features Method for Histopathological Image Classification. In: *Medical Image Computing ...* (2011). <http://www.springerlink.com/index/03815188524W8753.pdf>
- [140] RAZA S.H., Parry R M Sharma Y Chaudry Q Moffitt R A Young A N Wang M D.: Automated classification of renal cell carcinoma subtypes using bag-of-features. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*, 2010, 6749–6752
- [141] ROBERTS, Rosebud O. ; LYNCH, Charles F. ; JONES, Michael P. ; HART, Michael N.: Medulloblastoma: A Population-based Study of 532 Cases. In: *J Neuropath Exp Neur* 50 (1991), Nr. 2, 134–144. [http://journals.lww.com/jneuropath/Fulltext/1991/03000/Medulloblastoma\\_\\_A\\_Population\\_based\\_Study\\_of\\_532.5.aspx](http://journals.lww.com/jneuropath/Fulltext/1991/03000/Medulloblastoma__A_Population_based_Study_of_532.5.aspx)
- [142] ROUX, Ludovic ; RACOCEANU, Daniel ; LOMÉNIE, Nicolas ; KULIKOVA, Maria ; IRSHAD, Humayun ; KLOSSA, Jacques ; CAPRON, Frédérique ; GENESTIE, Catherine ; LE NAOUR, Gilles ; GURCAN, Metin: Mitosis detection in breast cancer histological images: An ICPR 2012 contest. In: *Journal of Pathology Informatics* 4 (2013), Nr. 1, S. 8
- [143] RUBIN, Adam I. ; CHEN, Elbert H. ; RATNER, Désirée: Basal-Cell Carcinoma. In: *New England Journal of Medicine* 353 (2005), Nr. 21, 2262–2269. <http://>

- [dx.doi.org/10.1056/NEJMra044151](http://dx.doi.org/10.1056/NEJMra044151). – DOI 10.1056/NEJMra044151. – PMID: 16306523
- [144] RUBINSTEIN, Reuven Y. ; KROESE, Dirk P.: *Simulation and the Monte Carlo method*. Bd. 707. John Wiley & Sons, 2011
- [145] SCHMIDHUBER, Jürgen: Deep learning in neural networks: An overview. In: *Neural Networks* 61 (2014), Oktober, 85–117. <http://dx.doi.org/10.1016/j.neunet.2014.09.003>. – DOI 10.1016/j.neunet.2014.09.003. – ISSN 08936080
- [146] SEMENZA, Gregg L.: Advances in cancer biology and therapy. In: *Journal of Molecular Medicine* 91 (2013), Nr. 4, 409–409. <http://dx.doi.org/10.1007/s00109-013-1024-2>. – DOI 10.1007/s00109-013-1024-2. – ISSN 0946–2716
- [147] SERTEL, O ; KONG, J ; SHIMADA, H ; CATALYUREK, U V. ; SALTZ, J.H. H. ; GURCAN, M.N. N. ; BOYER, K.L.: Computer-aided Prognosis of Neuroblastoma on Whole-slide Images: Classification of Stromal Development. In: *Pattern recognition* 42 (2009), Juni, Nr. 6, 1093–1103. <http://dx.doi.org/10.1016/j.patcog.2008.08.027>. – DOI 10.1016/j.patcog.2008.08.027. – ISSN 0031–3203
- [148] SICKLES, E A.: Breast cancer screening outcomes in women ages 40–49: clinical experience with service screening using modern mammography. In: *J Natl Cancer Inst Monogr* (1997), Nr. 22, 99–104. <http://www.biomedsearch.com/nih/Breast-cancer-screening-outcomes-in/9709284.html>. – ISSN 1052–6773
- [149] SOMMER, C. ; FIASCHI, L. ; HAMPRECHT, F. A. ; GERLICH, D.: Learning-based mitotic cell detection in histopathological images. In: *International Conference on Pattern Recognition (ICPR)*, 2012, S. 2306–2309
- [150] SPARKS, Rachel ; MADABHUSHI, Anant: Explicit shape descriptors: Novel morphologic features for histopathology classification. In: *Medical Image Analysis* 17 (2013), 2013 Dec, S. 997–1009. <http://dx.doi.org/10.1016/j.media.2013.06.002>. – DOI 10.1016/j.media.2013.06.002. – ISSN 1361–8423
- [151] SWEDLOW, Jason R. ; GOLDBERG, Ilya G. ; ELICEIRI, Kevin W.: Bioimage Informatics for Experimental Biology. In: *Annual review of biophysics* 38 (2009), Nr. 1, S. 327–346
- [152] TEZUKA, Shu: *Uniform random numbers: Theory and practice*. Bd. 315. Springer, 1995

- [153] TOŠIĆ, I ; FROSSARD, P: Dictionary learning. In: *Signal Processing Magazine, IEEE* (2011). [http://ieeexplore.ieee.org/xpls/abs/\\_all.jsp?arnumber=5714407](http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=5714407)
- [154] UHLEN, Mathias ; OKSVOLD, Per ; FAGERBERG, Linn ; LUNDBERG, Emma ; JONASSON, Kalle ; FORSBERG, Mattias ; ZWAHLEN, Martin ; KAMPF, Caroline ; WESTER, Kenneth ; HOBBER, Sophia ; WERNERUS, Henrik ; BJORLING, Lisa ; PONTEN, Fredrik: Towards a knowledge-based Human Protein Atlas. In: *Nat Biotech* 28 (2010), Dezember, Nr. 12, 1248–1250. <http://dx.doi.org/10.1038/nbt1210-1248>. – ISSN 1087–0156
- [155] VETA, M. ; PLUIM, J.P.W. ; DIEST, P.J. van ; VIERGEVER, M.A: Breast Cancer Histopathology Image Analysis: A Review. In: *Biomedical Engineering, IEEE Transactions on* 61 (2014), May, Nr. 5, S. 1400–1411. <http://dx.doi.org/10.1109/TBME.2014.2303852>. – DOI 10.1109/TBME.2014.2303852. – ISSN 0018–9294
- [156] VETA, Mitko ; DIEST, Paul J. ; WILLEMS, Stefan M. ; WANG, Haibo ; MADABHUSHI, Anant ; CRUZ-ROA, Angel ; GONZALEZ, Fabio ; LARSEN, Anders B. ; VESTERGAARD, Jacob S. ; DAHL, Anders B. ; CIREÇAN, Dan C. ; SCHMIDHUBER, Jürgen ; GIUSTI, Alessandro ; GAMBARDELLA, Luca M. ; TEK, F. B. ; WALTER, Thomas ; WANG, Ching-Wei ; KONDO, Satoshi ; MATUSZEWSKI, Bogdan J. ; PRECIOSO, Frederic ; SNELL, Violet ; KITTLER, Josef ; CAMPOS, Teofilo E. ; KHAN, Adnan M. ; RAJPOOT, Nasir M. ; ARKOUMANI, Evdokia ; LACLE, Miangela M. ; VIERGEVER, Max A. ; PLUIM, Josien P.: Assessment of algorithms for mitosis detection in breast cancer histopathology images. In: *Medical Image Analysis* (2014), –. <http://dx.doi.org/10.1016/j.media.2014.11.010>. – DOI 10.1016/j.media.2014.11.010
- [157] WANG, Chao ; PÉCOT, Thierry ; ZYNGER, Debra L. ; MACHIRAJU, Raghu ; SHAPIRO, Charles L. ; HUANG, Kun: Identifying survival associated morphological features of triple negative breast cancer using multiple datasets. In: *Journal of the American Medical Informatics Association* 20 (2013), April, Nr. 4, 1–8. <http://dx.doi.org/10.1136/amiajnl-2012-001538>. – DOI 10.1136/amiajnl-2012-001538. – ISSN 1527–974X
- [158] WANG, Haibo ; CRUZ-ROA, Angel ; BASAVANHALLY, Ajay ; GILMORE, Hannah ; SHIH, Natalie ; FELDMAN, Mike ; TOMASZEWSKI, John ; GONZALEZ, Fabio ; MADABHUSHI, Anant: Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection, 2014, 90410B-90410B-10

- [159] 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: *Journal of Medical Imaging* 1 (2014), Nr. 3, 034003. <http://dx.doi.org/10.1117/1.JMI.1.3.034003>. – DOI 10.1117/1.JMI.1.3.034003. ISBN 2329–4302
- [160] WEAVER, Donald L. ; KRAG, David N. ; MANNA, Edward A. ; ASHIKAGA, Taka ; HARLOW, Seth P. ; BAUER, Kenneth D.: Comparison of pathologist-detected and automated computer-assisted image analysis detected sentinel lymph node micrometastases in breast cancer. In: *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 16 (2003), November, Nr. 11, 1159–63. <http://dx.doi.org/10.1097/01.MP.0000092952.21794.AD>. – DOI 10.1097/01.MP.0000092952.21794.AD. – ISSN 0893–3952
- [161] WRIGHT, J. ; MA, Y. ; MAIRAL, J. ; SAPIRO, G. ; HUANG, T.S. ; YAN, S.: Sparse representation for computer vision and pattern recognition. In: *Proceedings of the IEEE* 98 (2010), Nr. 6, 1031–1044. [http://ieeexplore.ieee.org/xpls/abs/\\_all.jsp?arnumber=5456194](http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=5456194)