• U C •

FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Rui Miguel Morgado Venâncio

Micrometastasis detection guidance by whole-slide image texture analysis in colorectal lymph nodes correlated with QUS parameters.

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

Supervisors:
Prof. Dr. Daniel Racoceanu (UPMC, CNRS, INSERM, LIB, Paris, France)
Dr. Alain Coron (UPMC, CNRS, INSERM, LIB, Paris, France)

Coimbra, 2016

This work was developped in collaboration with:

Laboratoire d'Imagerie Biomédicale

Université Pierre et Marie Curie

Institut National de la Santé et de la Recherche Médicale

Centre National de la Recherche Scientifique

Resumo

O cancro é uma doença que afeta milhões por todo o mundo e uma identificação correta de gânglios linfáticos próximos do tumor primário, que contenham regiões metastáticas é de extrema importância para um correto gerenciamento dos pacientes. A avaliação histopatológica é o único método aceite para fazer essa identificação. Novas técnicas emergentes como os ultrassons quantitativos podem ajudar nessa identificação, detetando regiões metastáticas no gânglio linfático antes mesmo de o cortar. Propomos e avaliamos dois métodos para analisar e identificar automaticamente regiões suspeitas que contenham metástases em lâminas histopatológicas digitalizadas em alta resolução, guiando o patologista em direção às regiões suspeitas e classificando os gânglios como metastáticos ou não-metastáticos. O primeiro método, é um método convencional de análise de texturas e o segundo é baseado na aprendizagem profunda. Utilizando o método mais convencional participámos numa competição europeia chamada CAMELYON16. Esta competição tinha duas avaliações. Os métodos de textura utilizados foram as matrizes de coocorrência de níveis de cinzento e medidas de energia de texturas de Laws. Os parâmetros de textura serão utilizados para tentar encontrar relações entre os ultrassons quantitativos e a histopatologia. Para a aprendizagem profunda utilizamos uma rede bem documentada chamada VGG16. Imagens digitalizadas de lâminas histológicas de 44 gânglios foram utilizadas. Para avaliar os métodos foram desenhadas curvas ROC e F-Scores são calculados. Como resultados, obtivemos uma área sob a curva de 0.986 e um F-Score de 91.67 para o método mais convencional. Para a aprendizagem profunda obtivemos uma área sob a curva e um F-score igual a 1.0. Na competição ficámos em último numa avaliação e em penúltimo na outra. Para finalizar, não foi possível encontrar nenhuma correlação entre os ultrassons e a histologia.

Abstract

Cancer is a disease that affects millions worldwide and accurate determination of whether lymph nodes (LNs) near the primary tumor contain metastatic foci is of critical importance for proper patient management. Histopathological evaluation is the only accepted method to make that determination. New emerging techniques like quantitative ultrasound (QUS) may help in the determination by detect metastatic regions in the LN before cutting it. We propose and evaluate two methods to automatically analyze and identify suspicious regions for metastatic foci in high-resolution digitized histopathological slides (whole-slide images (WSI)) to helping the guidance of the pathologist towards cancer-suspicious regions and to classify LNs as metastatic or non-metastatic. The first method is a conventional texture-based method and the second one is based in deep convolutional neural networks (DCNNs). We have participated in the CAMELYON16 challenge with the conventional method. The texture methods used are based on gray-level co-occurrence matrices (GLCM) and Laws' energy texture measures, which parameters will be used for find correlations with the QUS. As DCNN we used a known network called VGG16. Whole slide images (WSIs) of 44 lymph nodes (LNs) were used. For evaluate both methods Receiver Operating Characteristic (ROC) curves were drawn. For the most conventional method we obtained an Area Under the Curve (AUC) of 0.986 and a F-Score of 91.67. For the CNN based method we obtained an AUC and a F-Score of 1.0. The challenge had 2 evaluations, and we came last in one and second-to-last in the second. We could not find any correlation between the ultrasounds and the histology

List of Acronyms

1D One-Dimensional.
2D Two-Dimensional.
3D Three-dimensional.

AUC Area Under the Curve.

BBMR Boosted Bayesian Multiresolution.

CAD Computer-aided detecion.
CFS Correlation-based Feature Selection.
CRC Colorectal Cancer.
CT Computed Tomography.

DCNN Deep Convolutional Neural Network.

FCFS Fast correlation-based feature Selection.
FCNN Fully Convolutional Neural Network.
FN False Negative.
FP False Positive.
FROC Free-Response Receiver Operating Characteristic.

GLCM Gray-Level Co-occurrence Matrices.

H&E Hematoxylin and Eosin.
HF High Frequency.
HPF High-Power Field.

LBP Local Binary Patterns.
LN Lymph Node.

MBF Markov Blanket Filter.

QUS Quantitative Ultrasound.

RBF Radial Basis Function.
ROC Receiver Operating Characteristic.
ROI Region of Interest.
RR Round-Robin.

SBE Sequential Backward Elimination.
SFS Sequential Forward Selection.
SVM Support Vector Machine.

TEM Texture Energy Measurements.
TN True Negative.
TP True Positive.

US Ultrasound.

WSI Whole-Slide Image.

List of Figures

List of Tables

Contents

1

Introduction

Each year, millions are affected worldwide by cancer and it is of critical importance to detect it at early stages in order to provide high chances of survival. One of the main concerns is when the disease spreads, through blood or lymphatic systems, from the place where it started to another place in the body (forming metastases) in a process called metastasis [3]. In many types of cancers, an accurate determination of whether Lymph Nodes (LNs) contain metastatic foci (micrometastases and isolated tumor cells) is vital to determine disease stage and proper patient management.

1.1 Lymphatic System and Cancer

In the presence of harmful agents (external and/or internal) our body defends itself via an immune response. Not being exclusively responsible, the lymphatic system is important in this type of response due to its different constituents and its functions [1]. This system is mainly composed by bone marrow, thymus, spleen, lymph nodes and different types of lymph vessels (Figure 1.1). Lymph vessels forms a network that transports lymph around the body. Lymph is composed by the part of the fluid and the substances (viruses, bacteria, antigens, lymphocytes) present in the extracellular space of the connective tissue [1, 4] that gain entry in the lymphatic capillaries. While blood vessels carry oxygen, nutrients, wasting products and carbon dioxide, LNs work as filters for the harmful agents/substances. Sometimes, LNs swell, which is a condition that may be caused by an infection, injury or cancer [4].

When cancer spreads from the original tumor to distant sites, especially in breast, colorectal or gastric cancer freed cancer cells pass mainly through lymphatic vessels at tumor site and spread through lymphatic system. This is a highly inefficient process. Only 0.01% of tumor cells that enter in bloodstream/lymph develop into metastases, because they die or they are killed [5]. The LN where the tumor cells

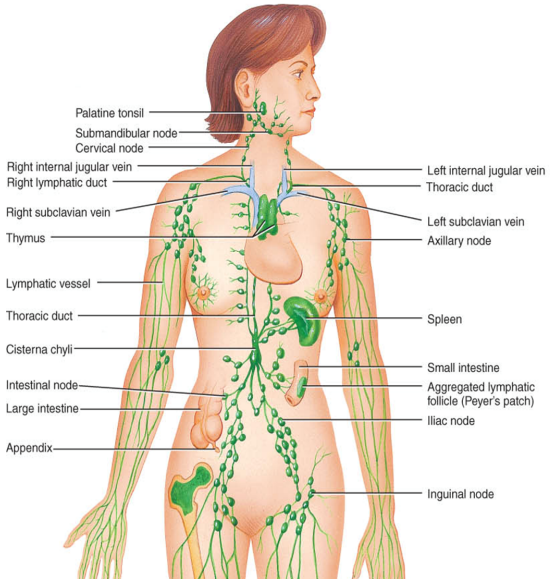


Figure 1.1: Lymphatic system and its components [www.medicinenet.com]

settle is referred as sentinel LN and normally is a LN near the original tumor [6].

Normally, the LNs near a primary cancer are removed when the surgeon removes the cancer to look for cancer signs in them [4]. The way a LNs is examined to look for cancer regions is explained later.

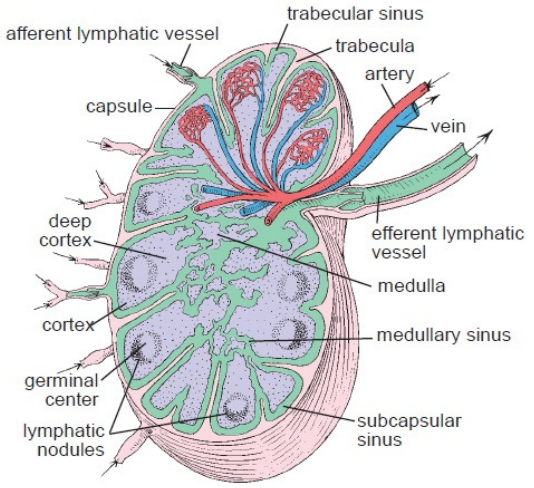


Figure 1.2: General structure of a lymph node [1]

1.2 Lymph Node Anatomy

A LN is a small bean-shaped organ, whose dimension varies from 1mm to 2cm. They are distributed all over the body, but in certain regions such as the axilla, groin, and mesenteries, their density is higher (Figure 1.1). In Figure 1.2, the general structure of a LN is presented.

Lymph fluid enters and exits the LN via the afferent (in) and efferent (out) lymphatic vessels. The anatomy of the LN can be divided into different structures and adjacent regions: the cortex as a more external region and the medulla, the inner region of the LN; the capsule, which covers the organ; the trabeculae which is composed by connective tissue that is localized between the capsule and the lymphatic sinus (trabecular and subcapsular) of the organ; finally, the reticular tissue, which is composed by reticular cells and fibers forming a supporting meshwork. In addition to the reticular cells, this meshwork is composed by dendritic, macrophages and follicular dendritic cells [1].

1.3 Methods for Metastases Detection

After the surgical removal of the LN due to suspicious of metastasis, it is of utmost importance to confirm the diagnosis. The gold standard of this diagnosis is the analysis of the dissected LN by a pathologist. However, there is an alternative and interesting technique, Ultrasound (US).

1.3.1 Histology

The gold standard to determine the presence of cancerous tissue is the microscopic histologic evaluation of dissected LNs by a pathologist. For a typical histological evaluation of colorectal cancer, only a section from each LN is cut, which is a small portion of the LN to be evaluated [7]. Evaluating more slices (2 or 3) is not the current histological practice as it would be an exhausting and time-consuming task for the pathologist. However, it would reduce the number of false negative LN [8].

1.3.2 3D Quantitative Ultrasound: An Emerging Technique for Metastasis Detection

The ultrasound can solve the problem of analysis time because it can evaluate the entire LN. Saegusa-Beercroft *et al.* [9] tried to develop a method using High Frequency (HF) Quantitative Ultrasound (QUS) for an entire-volume LN examination. Three-dimensional (3D) ultrasound scans have the ability of acquire data from the whole LN. If this data is combined with the QUS parameters, it is possible to QUS methods to evaluate an entire LN, detecting both micro and macrometastases. With this information it is possible to guide the pathologist to cut the LN in a specific zone, to guarantee that it does not miss any metastatic tissue.

1.4 Objective

As histopathology is the gold standard for metastasis detection a first approach consists in setting up a texture-based Computer-aided detecion (CAD)[1] system to identify metastatic and non-metastatic colorectal LNs in high-resolution digitized histological slides (WSI). Then those histologic features that were valuable for this CAD system were correlated with some QUS parameters that were also shown to be of high interest for metastatic detection with QUS [10]. By correlating them, we wished to know whether one could identify a linear relation between at least some of the histologic and US parameters.

In order to test the approach on another histologic database, a slightly modified CAD system was enrolled to CAMELYON16 challenge organized during the IEEE ISBI conference to detect metastasis in lymph nodes of breast cancer patients.

Finally, a second CAD system based on Deep Convolutional Neural Network (DCNN) was developed and evaluated to detect suspicious regions in LNs.

[1]The majority of authors refer to computer-aided diagnosis rather than computer-aided detection. Due to the nature of the problem we choose the latter one.

2

Data

2.1 Ultrasounds Data on Colorectal Lymph Nodes

A total of 112 lymph nodes from 77 patients with a proven colorectal or gastric cancer were dissected and scanned at Kuakini Medical Center in Honolulu, HI, USA. Ninety-two of them were free of metastases, while the remaining 20 were almost filled with metastases. Prior to histopathological evaluation, the LN were scanned in 3D with US. Radio-frequency signals were acquired using a focused, single-element transducer (PI30-2-R0.50IN, Olympus NDT, Waltham, MA, USA) with a center frequency of 25.6 MHz. The envelope statistics of the backscattered signal was modelled and eight 3D QUS parameters were obtained [10]. From the eight only two of them were used. The Effective Scatterer Size ($MeanR$) and the Acoustic concentration ($MeanC$). For each LN, the mean value of these two parameters were specially computed for this work.

2.2 Whole-Slide Images of Colorectal Lymph Nodes

All the 112 lymph nodes used to obtain the QUS parameters were then prepared for histology as described in Mamou *et al.* [11] and Hematoxylin and Eosin (H&E) stained slides were obtained every 50 μm (or 100 μm for large LNs). Those slides were as much as possible perpendicular to the axis of the transducer. Among the 112 LNs, 44 were scanned using a Hamamatsu NanoZoomer Whole Slide Scanner at 20x magnification (i.e., approximately 0.45 μm per pixel). A total of 270 slides were acquired. The remaining LNs are scanned each time one can get an access to

such a device. The size of the WSIs varies from 10496 to 51712 pixels in height and 11520 to 99840 pixels in width.

This group of 44 LNs was divided into two groups: a 1^{st} group of 20 LNs (15 cancerous and 5 non-cancerous totaling 160 slides) to train and validate both methods, and another for testing with 24 LNs (13 cancerous and 11 non-cancerous totaling 110 slides). On 19 slides of the first group, a pathologist drew green (respectively blue) rectangles on cancerous (respectively cancer free) regions (Figure 2.1), using a software called NDP.view 2. Each annotation is represented by the coordinates of each one of the 4 corners of the rectangle, and are stored in a ndpa file. These regions were divided into non-overlapping sub-images that will be referred to as HPFs.

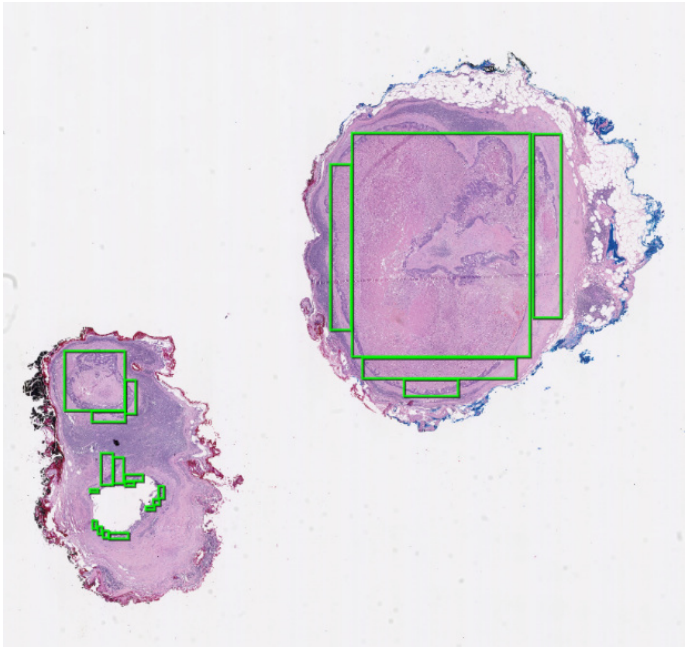


Figure 2.1: Cancerous annotations drawn by the pathologist

In Figure 2.2, a black and white low resolution photograph of a physical glass slide is shown. One can read the slide label on the left side of the slide and one can see two consecutive thin slices of LN tissue on the right. On the low resolution image acquired with the NanoZoomer slide scanner, some color marks are represented. In red, we have the border of the physical slide and its center, C_r. The yellow rectangular region represents the rectangular region that was truly scanned at a high resolution. The yellow point represents the center C_y of this high resolution region.

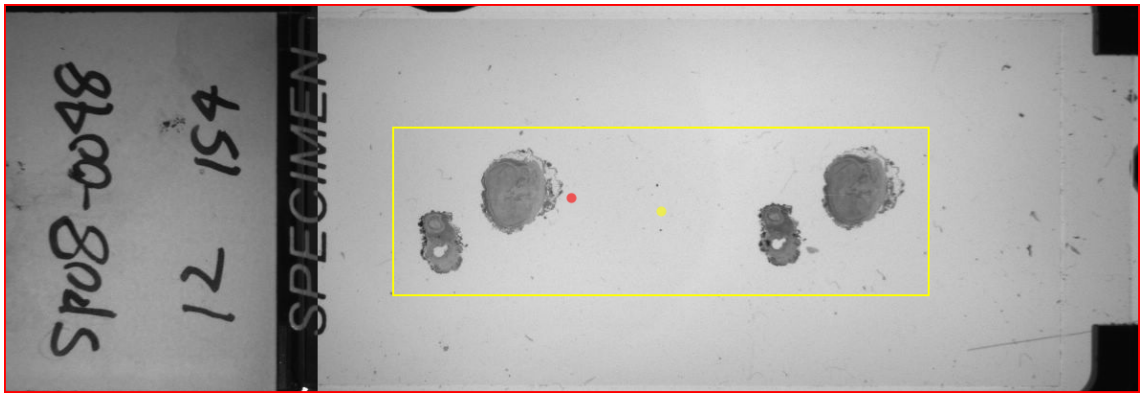


Figure 2.2: Black and white photography of a slide ready to be analyzed in the microscope or to be scanned. This type of image is designed as "macro"

2.3 Whole-Slide Images of Breast Lymph Nodes (CAMELYON16)

The database was provided by the organizer of the CAMELYON16 challenge organized during the IEEE ISBI conference. It contains 400 WSIs of sentinel LNs from patients with breast cancer. This database is composed by two independent datasets collected in Radboud University Medical Center and the University Medical Center Utrecht. Its WSIs size vary from 28672 to 221696 pixels in height and 61440 to 221184 pixels in width.

The database was divided in two training datasets and one test dataset. The first training dataset consists in 170 WSIs (100 negatives and 70 positives) and the second dataset with 100 WSIs (60 negative and 40 positive). The test dataset contains 130 WSIs. The resolution of each WSI is 0.225 μm per pixel.

Instead of having rectangular annotations like the colorectal database, this database has another type of annotations. Each positive WSI has a binary image identifying the metastatic regions, that are also divided in non-overlapping HPFs.

2.4 Whole-Slide Image Files and Their Pyramidal Structure

Each WSI is organized in a pyramid structure, where each level of the pyramid represents a different magnification level. The lower the level the pyramid the higher the magnification and consequently more details presented in the image. With

(a) **(b)** **(c)**

Figure 2.3: Example of CAMELYON16 WSI (a), respective annotation (b) and the superposition of both (c)

the increasing of the pyramid levels the details in the images decrease due to the decreasing of magnification levels. Table 2.1, shows the different dimensions and magnification levels of the same image at different levels of the pyramid. Matlab (The MathWorks, Inc., Natick, Massachusetts, United States) and OpenSlide [12] were used to play around with the WSIs.

Table 2.1: Example of dimensions and magnifications at different levels in pyramid structure of a WSI. The base-level of the pyramid is the level-0 while the top-level is the level-8.

Level	Dimension ($H \times W$)	MAgnification Level
0	37632×74880	20
1	18816×37440	10
2	9408×18720	5
3	4704×9360	2.5
4	2352×4680	1.25
5	1176×2340	0.625
6	588×1170	0.313
7	294×585	0.156
8	147×292	0.078

3

State of the Art and Methods

In this chapter all the methods that were used in this work will be explained. We will start to explain how HPFs are extracted from the WSIs. We present you some of the works found in the literature, the texture methods as well the machine learning algorithm we used. To finish we describe you step-by-step our approaches and the evaluation used in this work.

3.1 HPF Extraction

We built a modifiable general application to select and extract valid HPFs from both databases. First, a description of how the HPFs are extracted from the annotated and non-annotated WSIs of our database is presented and then which modifications were introduced in order to be able to use the application with the challenge database.

3.1.1 The Colorectal Database

The coordinates stored in the ndpa file do not represent specific locations (in pixels) in WSI, but the distance, in nanometers (nm), from the corners to the center of the slide, being necessary to convert these distances in pixel locations in the WSI.

3.1.1.1 From nanometers to pixels

As we said before, the coordinates of the annotations are in nm and represent the distance from any point inside the yellow rectangle to the red point (Figure 2.2). Using the Openslide library we can obtain the distance from C_r to C_y, we call this

the offset from the center. In order to obtain the coordinates relatively to C_y, we subtract the offset to the original coordinates.

In Matlab, the origin of an image is the left-upper corner. We need to do some calculations to obtain the coordinates in pixels and relatively to the origin, viewed by Matlab. Again, using Openslide, we obtain the resolution, in micrometers per pixel (μm/pixel), of the WSI and the its size (in pixels). Finally, using the equation 3.1, for both directions x and y, we obtain the coordinates in pixels

$$C_p = \frac{(10^{-3} \times C_{nm})_{\mu m}}{R_{\mu m/p}} + \frac{S_{WSI_p}}{2}, \tag{3.1}$$

where C_p is the coordinates in pixels, C_nm the coordinates in nm, $R_{\mu m/p}$ is the resolution of the WSI in μm/pixel and finally S_{WSI_p} is the size (width or height) of the WSI in pixels.

3.1.1.2 HPF of annotated WSI

Once we have the coordinates in pixels, we can start the process of getting the HPFs. As stated above, images at high magnification are larger, so in this step we performed the necessary pre-processing and calculations in images at low magnification (0.625) so at a level 5 in the pyramid (see Table 2.1).

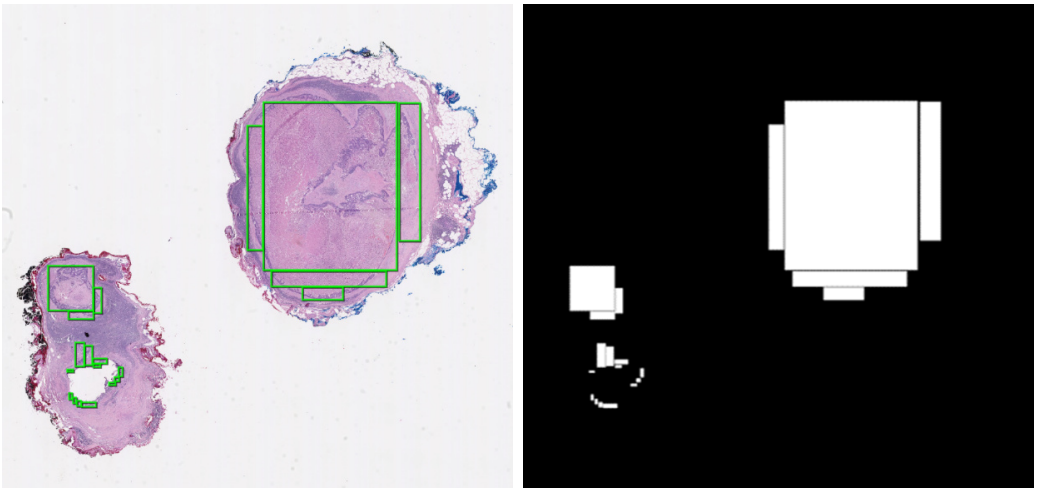
We need to calculate the coordinates of the annotations at level 5, using the coordinates at level 0, as well as the resolution at level 0. The first thing to do is to calculate the resolution of the image at level 5, using equation 3.2

$$R_{L\#} = 2^{\#} \times R_{L0}, \tag{3.2}$$

where # is the level desired (# can be any value from 1 to the maximum number of levels of that specific slide), $R_{L\#}$ is the resolution at level # and R_L0 is the resolution at level 0. After that, a binary mask is created (Figure 3.1b).

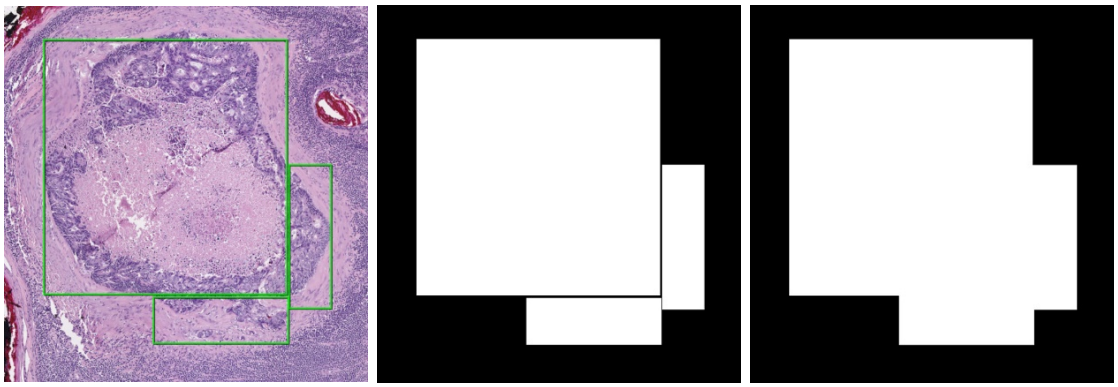
In the Figure 3.1b, it is possible to see some small gaps between annotations that are very closed and of the same type. In order to eliminate them the annotations were merged using mathematical morphology obtaining that way the final mask (Figure 3.2c).

After merging the annotaions it is time to proceed to the division of the mask in HPFs. We get the x and y-coordinates of the left-most and upper-most points of the mask and the point composed by the minimum x and y values between these



(a) **(b)**

Figure 3.1: Annotations (a) and respective binary mask (b)



(a) **(b)** **(c)**

Figure 3.2: Original image with the annotations (a) and the respective masks before (b) and after (c) the mathematical morphology operation

two coordinates is used as starting point. In Figure 3.2, the starting point coincides with left-upper corner of the mask, but this optimal case not always happens. With the starting point determined, the number of HPFs to be extracted is calculated by dividing the width and the height of the annotation by the width and height of the HPFs, rounding the result to the nearest integer towards infinity. Now, from the starting point a grid is created and the HPF extracted. In Figure 3.3, we can visualize the Figure 3.2a divided in HPFs of size 1000×1000 pixels. It is important to mention that not all HPFs are valid. Only those that does not have background (black regions) in it are considered valid. In Figure 3.3, we can see that only two out of nine are saved.

3.1.1.3 HPF extraction from slides without annotations

When evaluating the methods, WSIs do not have any annotations, HPFs have to been automatically created to identify LN sections, and mask fatty tissue, background and tears. It starts by applying Otsu's [13, 14] method to the intensity of the high resolution WSI to find the global image threshold allowing the conversion of the WSIs in a binary image (Figure 3.4).

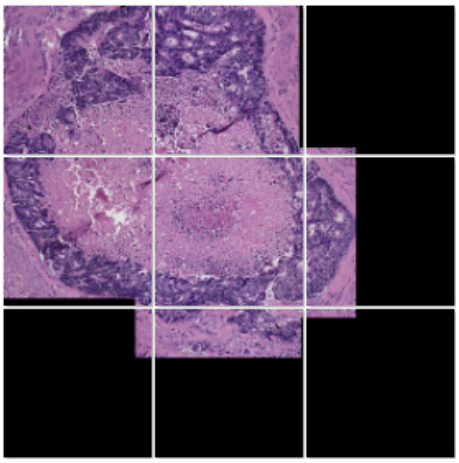
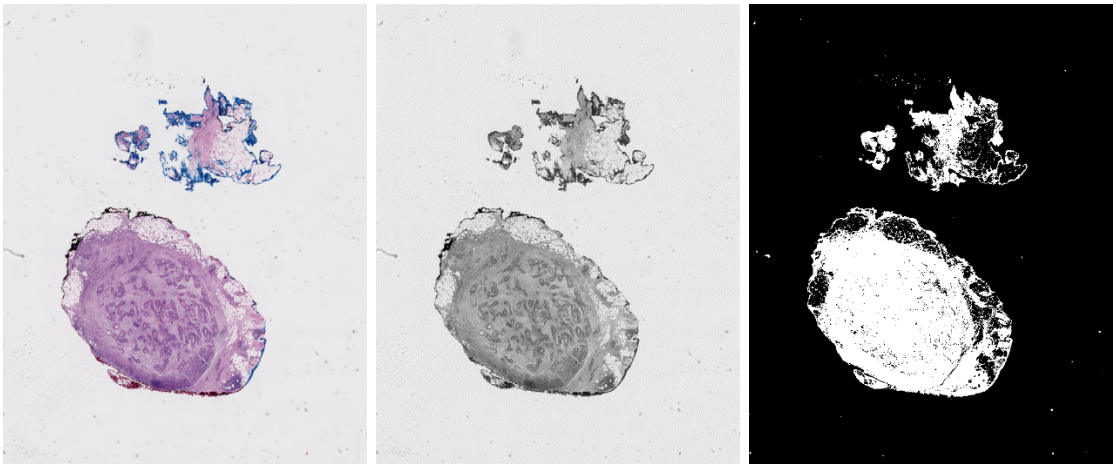


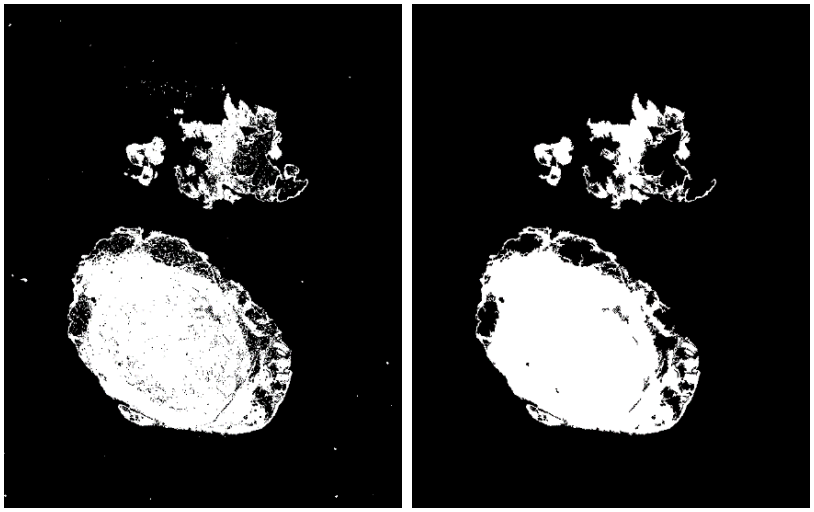
Figure 3.3: Annotation divided in HPFs



(a) **(b)** **(c)**

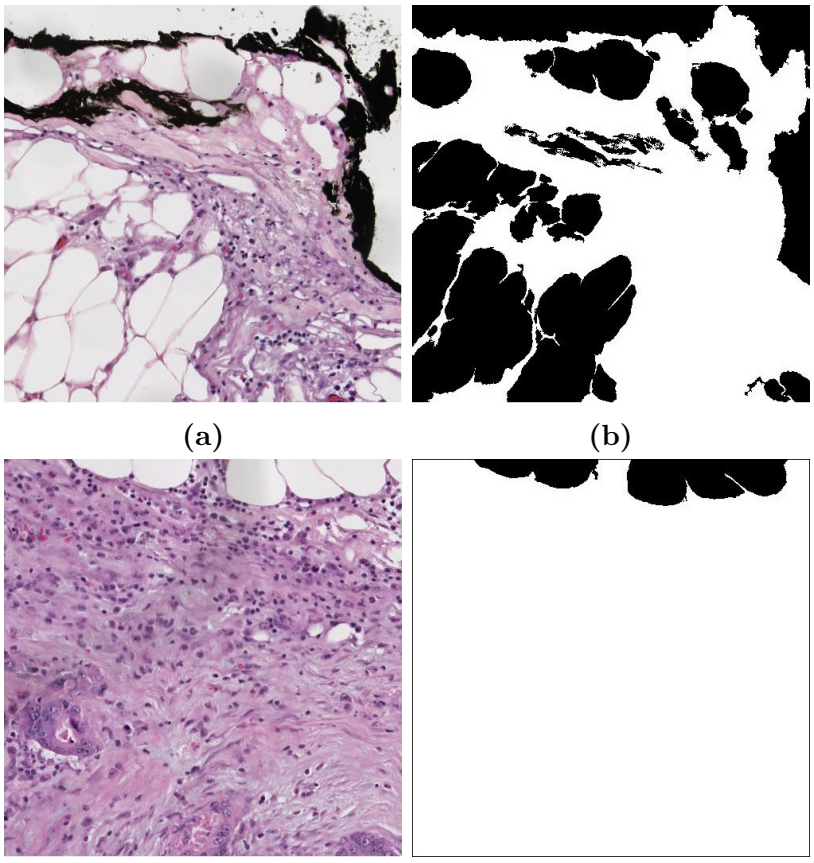
Figure 3.4: (a) and (b) represent the WSI (RGB and intesity images) before the application of Otsu's method, (c) is the result of the application of the method.

The binary image has some unwanted regions in the background that need to be remove and some wanted regions in the LN section that need to be added. The removal and addition processes are achieved by using mathematical morphology, allowing the obtention of the final mask (Figure 3.5b).



(a) **(b)**

Figure 3.5: Before (a) and after (b) the removal and addition steps.



(a) **(b)**

(c) **(d)**

Figure 3.6: Two examples of valid and invalid HPFs when annotations are not present. At (a) and (b) is the example of an invalid HPF and its respective mask. The area of background (black) in this specific HPF is around 46% of the total area of the HPF. At (c) and (d) is the example of a valid HPF and its respective mask. In this case, the area occupied by the background is around 5% of the total area of the HPF

Now, the process follows a similar path as the one described in Section 3.1.1.2. The Region of Interests (ROIs) shown in Figure 3.5b are now divided in non-overlapping consecutive HPFs. In this particular case there are 3 different regions, so for each region the x and y-coordinates of the left-most and upper-most points of the mask are found. The point formed by the minimum x and y values between the previously found coordinates is used as starting point. Once the starting point is found, the process of extracting the HPFs is similar of that with annotations with one exception, the definition of a valid HPF. An HPF is valid if the area of the background pixels is lower than 10% of the size of the HPF (Figure 3.6). In Figure 3.7 is possible to the result of dividing a LN in HPFs.

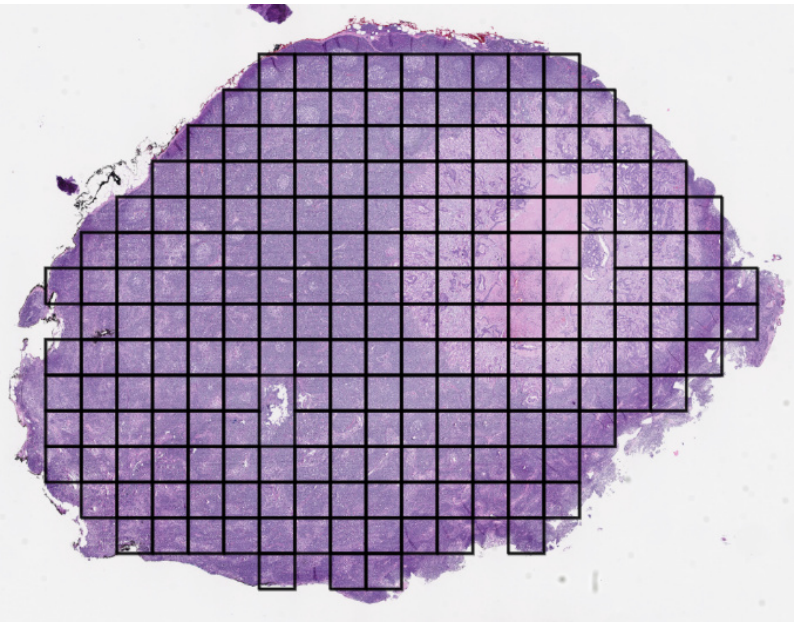


Figure 3.7: WSI divided in HPFs

3.1.2 CAMELYON16 database

The general process of extracting HPFs is the same as described in Section 3.1.1.3, with one difference. Instead of using Otsu's method to convert the grayscale image to binary image, a K-means method is used [14].

With the LN divided in HPFs as in Figure 3.7. If the WSI belongs to a non-metastatic LN (or does not have a binary mask), all the HPFs are treated as non-metastatic. If the LN is metastatic, we use the binary mask that identifies the metastatic regions and the HPF that at least 90% of its area lays in those metastatic regions are treated as metastatic HPFs, while the remaining as non-metastatic.

3.2 Texture analysis in histology

3.2.1 State of the art

3.2.1.1 Conventional Method

Kather *et al.* [15] used textural descriptors to perform a multiclass texture separation in 5000 WSI patches of size 150 × 150 pixels of Colorectal Cancer (CRC) histological images containing 8 different types of tissue. They used six different textural descriptors: Low-order and higher-order histogram features, Local Binary Patterns (LBP), GLCM, Gabor Filters, Perception-like features and a combination of all texture descriptors. As results, they obtained 98,6% accuracy for tumor-stroma separation while for multiclass tissue separation 87.4% (a new standard). Doyle *et al.* [16] presented a Boosted Bayesian Multiresolution (BBMR) system to identify prostate cancer regions in WSI of biopsies of prostate tissue. From the original images they obtained different images at different resolution levels (low, medium and high). The cancerous regions detected at low resolution were examined at higher resolution, increasing that way the analysis efficiency of large images. For each resolution level a total of 900 first-order statistical (average, median, etc.), second-order co-occurrence (GLCM) and steerable filter (Gabor filters) features were obtained and reduced to 10 by applying an AdaBoost [17] method. Using a Bayesian classifier, area under the receiver operating characteristic curve (AUC) values of 0.84, 0.83 and 0.76 were obtained for the lowest, intermediate and highest resolution levels. Bouatmane *et al.* [18] proposed an automatic classification system to diagnose prostate cancer in images obtained from multispectral data of prostate needle biopsies using textural (GLCM) and some structural features. For the diagnosis of prostate cancer, it is necessary to accurately identify four types of structures: stroma, benign prostatic hyperplasia, prostatic intraepithelial neoplasia and prostatic carcinoma. In order to use a binary classifier to handle multi-class problems a Round-Robin (RR) Sequential Forward Selection (SFS) (RR-SFS) algorithm was used, where RR is a classification technique that divide the multiclass problem into simpler binary classification problems where each classifier is a sequential forward selection/nearest neighbor (SFS/1NN) classifier. Using this technique, the authors state an accuracy of 0.999. Nguyen *et al.* [19] proposed an automatic cancer detection system where the images were divided into a grid of patches and from each patch cytological and textural features were extracted. The textural features extracted include first-order statistics like mean, standard deviation, etc., second-order statistical features

obtained from GLCM and Gabor filter-based texture features. Using a Support Vector Machine (SVM) classifier with Radial Basis Function (RBF) kernel they obtained a true positive rate of 0.78 and a false positive rate of 0.06. Elnemer [20] developed an automatic CAD system for early detection of lung cancer and lung water on Computed Tomography (CT) images using Laws' energy measures. First, some preprocessing techniques are performed in the images. In the preprocessed images lung regions are segmented and Laws' texture features extracted. Finally, using a t-test, the features that best discriminated cancer, water lung and normal cases were selected. Rachidi [21] apply Laws' masks on high-resolution digital radiographs in an attempt to describe structural variations of trabecular bone. He showed that Laws' masks could become a promising technique to determine osteoporosis fracture risk.

3.2.1.2 Deep Convolutional Neural Networks

With a first success in the 90s [22], the DCNNs have become nowadays state of the art methods in image classification [23], segmentation [24] and in dimensionality reduction [25]. Their introduction in digital histopathology was made possible in part through challenges: Gland Challenge Contest Segmentation MICCAI in 2015 [26], ICPR 2012 competition mitosis detection, etc. These challenges were made public databases annotated by experts. This allowed a good comparison of different approaches on histopathology images as mentioned in [27].

In the last five years, have emerged three major DCNNs use cases/domains in digital histopathology. One of the first was mitoses detection on histological images of breasts, by Ciresan *et al.* [28]. The authors classify each pixel of the image using a patch centered in the latter. Using a data augmentation, they managed to train the network from scratch. Su *et al.* [29] combined a sparse reconstruction and has stacked denoising auto-encoders to perform cell detection and segmentation in lung cancer and brain tumors. Lately, Sirinukunwattana *et al.* [30] propose a spatially constrained convolutional neural network (CNN-SC) to perform nucleus detection.

In segmentation, Ronneberger *et al.* [31] proposed a Fully Convolutional Neural Network (FCNN) called U-Net. It is composed of a convolutional block followed by a deconvolutional one. A set of rotations and elastic deformations has been used to increase the training set. Chen *et al.* [32] also used FCNN and won the glands segmentation challenge [26]. As against to avoid a step of increasing the learning base, they made recourse to transfer learning. The network was trained from weight

learned on Pascal VOC 2012.

In classification, Spanhol *et al.* [33] used a DCNN to classify breast cancer histopathological images. A set of patches is extracted from the basic images to train the network. Litjens *et al.* [34] went further by proposing a classification method of WSIs.

The more example we use in the training the better accuracy we get. That explain the necessity of data augmentation. But, some approaches [24, 31, 32] show that the use of a pretrained network can overcome this obstacle. That explain our proposed approach.

3.2.2 Methods Selected for Features Extraction

3.2.2.1 Gray-Level Co-occurrence Matricess and their Associated Features

This method, as the title suggests, uses co-occurrence matrices built from the pixel gray values of the image. These matrices are squared and the number of rows and columns is equal to the gray levels of the image. A grayscale image, contains 256 gray levels, so it is usual to requantize them to a smaller value, in our case 8. Each element in the matrix is represented as $P(i, j|d, \theta)$, where i and j are the position in the matrix, d the interpixel distance and θ the direction of the analysis [35]. The value P is the number of transitions between the positions i and j, separated by d pixels in the direction defined by θ. A co-occurrence matrix can be symmetric or non-symmetric. In our case we used non-symmetric matrices.

An illustrative GLCM example is shown in Figure 3.8. Four different angles were used $(0°, 45°, 90°, 135°)$ and a interpixel distance of one was chosen. For example, to get the value of $P(0, 0|1, 0)$ one counts in the image how many time 0-value pixels have another 0-value pixel as its right neighbor (East Direction). For $P(3, 2|1, 135)$, one count how many 3-value pixels have a 2-value pixel as its North West neighbor.

In this work, interpixel distances between 1 and 10 combined with four angles (i.e.,$0°, 45°, 90°, 135°$) were used yielding a total of 40 different co-occurrence matrices.

Directly from the co-occurrence matrices features can be extracted. Thirteen features from the 14 proposed by Haralick *et al.* [36] were used. (For readability

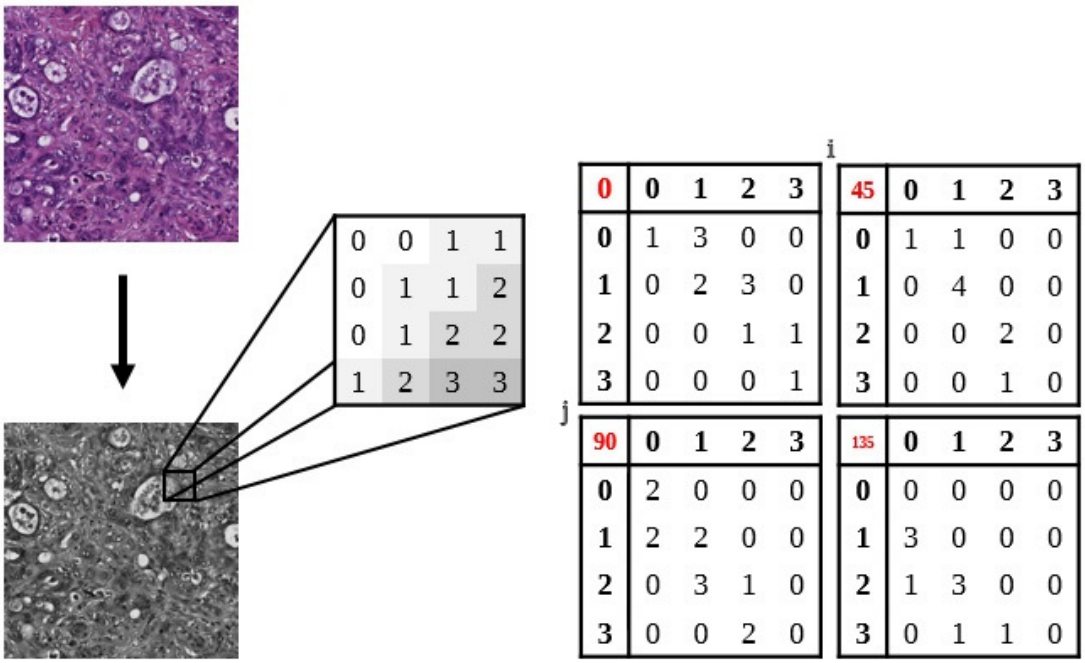


Figure 3.8: A GLCM example. The interpixel distance d, is equal 1 and the angles θ are 0, 45, 90 and 135 degrees.

distance and angle were omitted from the equations, but these thirteen features were computed for all 40 co-occurrence matrices.)

Angular Second Moment $f_1 = \sum_i \sum_j \{p(i,j)\}^2$

Contrast $f_2 = \sum_{n=0}^{N_g-1} n^2 \{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)\}, |i-j| = n$

Correlation $f_3 = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$

Sum of Squares: Variance $f_4 = \sum_i \sum_j (i-\mu)^2 p(i,j)$

Inverse Difference Moment $f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$

Sum Average $f_6 = \sum_{i=2}^{2Ng} i p_{x+y}(i)$

Sum Variance $f_7 = \sum_{i=2}^{2Ng} (i - f_9)^2 p_{x+y}(i)$

Sum Entropy $f_8 = -\sum_{i=2}^{2Ng} p_{x+y}(i) \log\{p_{x+y}(i)\}$

Entropy $f_9 = -\sum_i \sum_j p(i,j) \log(p(i,j))$

Difference Variance $f_{10} = variance of p_{x-y}$

Difference Entropy $f_{11} = -\sum_{i=0}^{Ng-1} p_{x-y}(i) \log\{p_{x-y}(i)\}$

Information Measures of Correlation

$f_{12} = \frac{HXY - HXY1}{\max\{HX, HY\}}$

$f_{13} = \sqrt{(1 - e^{-2[HXY2 - HXY]})}$

where R is the total number of pairs in the matrix, Ng is the number of gray levels in the image, μ_x, μ_y, σ_x, σ_y are the mean and the standard deviation of p_x and p_y, and

$$p(i,j) = \frac{P(i,j|d,\theta)}{R} \tag{3.3}$$

$$p_x(i) = \sum_{j=1}^{Ng} p(i,j) \quad p_y(j) = \sum_{i=1}^{Ng} p(i,j) \tag{3.4}$$

$$p_{x+y}(k) = \sum_{i=1}^{Ng}\sum_{j=1}^{Ng} p(i,j), i+j = k, k = 2,3,...,2Ng \tag{3.5}$$

$$p_{x-y}(k) = \sum_{i=1}^{Ng}\sum_{j=1}^{Ng} p(i,j), |i-j| = k, k = 2,3,...,Ng-1 \tag{3.6}$$

$$HXY = -\sum_i \sum_j p(i,j)\log(p(i,j)) \tag{3.7}$$

$$HXY1 = -\sum_i \sum_j p(i,j)\log\{p_x(i)p_y(j)\} \tag{3.8}$$

$$HXY2 = -\sum_i \sum_j p_x(i)p_y(j)\log\{p_x(i)p_y(j)\} \tag{3.9}$$

3.2.2.2 Laws' Filters and their Associated Features

Laws texture energy measures consist on sets of One-Dimensional (1D) filters (14 in total) at different scales: 3, 5, 7. These 1D masks were created by Kenneth Laws [37] and those of size 3 and 5 are presented below:

$$\begin{aligned}
\mathbf{L3} &= \begin{matrix} 1 & 2 & 1 \end{matrix} \\
\mathbf{E3} &= \begin{matrix} -1 & 0 & 1 \end{matrix} \\
\mathbf{S3} &= \begin{matrix} -1 & 2 & 1 \end{matrix} \\
\mathbf{L5} &= \begin{matrix} 1 & 4 & 6 & 4 & 1 \end{matrix} \\
\mathbf{E5} &= \begin{matrix} -1 & -2 & 0 & 2 & 1 \end{matrix} \\
\mathbf{S5} &= \begin{matrix} -1 & 0 & 2 & 0 & -1 \end{matrix} \\
\mathbf{W5} &= \begin{matrix} -1 & 2 & 0 & -2 & 1 \end{matrix} \\
\mathbf{R5} &= \begin{matrix} 1 & -4 & 6 & -4 & 1 \end{matrix}
\end{aligned}$$

The masks of size 3 represent the basis for the other larger masks. Masks of size 5 can

be obtained by convolving two 3-sized masks. To obtain 7-sized masks a convolution between a 3-sized and a 5-sized mask is required. The first letter in the mask names stands for either Level, Edge, Spot, Wave, Ripple or Oscillation. These horizontal 1D masks are convolved with the same vertical masks to obtain Two-Dimensional (2D) masks that are convolved with the images to extract individual structural elements. For example, combining a $E3$ horizontal mask with a $S3$ vertical mask, results in a 2D mask E_3S_3, that convolved with an image gives the edges in the horizontal direction and the spots in the vertical direction.

In this work, the procedure described in [21] was applied: images were convolved with the 2D masks. For example, convolving image $I(i,j)$, with the mask E_3S_3, results in image $I_{E_3S_3}(i,j)$.

$$I_{E_3S_3}(i,j) = I(i,j) \otimes E_3S_3 \tag{3.10}$$

In order to normalize the contrast of the resulting images, all the resulting images are divided by $I_{L_3L_3}(i,j)$, resulting in a normalized image $NI_{E_3S_3}(i,j)$

$$NI_{E_3S_3}(i,j) = \frac{I_{E_3S_3}(i,j)}{(I_{L_3L_3}(i,j)} \tag{3.11}$$

Finally, the normalized images are passed to a 15×15 Texture Energy Measurements (TEM) filters to create texture energy maps representing the application of the 2D masks to the input image.

$$TEM_{E_3S_3}(i,j) = \sum_{u=-7}^{7} \sum_{v=-7}^{7} [NI_{E_3S_3}(i+u, j+v)] \tag{3.12}$$

When the same 1D masks are convolved, changing their orientations, such as for example the masks E_3S_3 and S_3E_3, a rotationally invariant (only for $\theta = 90°$) texture energy measurement denoted by RT needs to be created. For the proposed example

$$RT_{E_3S_3} = \frac{TEM_{E_3S_3} + TEM_{S_3E_3}}{2}. \tag{3.13}$$

Convolving the 1D masks between them, it is possible to obtain 5, 14 and 20 RTs for the three different sizes (3, 5 and 7 respectively), totaling 39 images from where features can be calculated.

Once the RTs are created, 5 statistical features can be extracted from them. These

features are described in [21] and they are

$$\text{Mean} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(RT_{ij})}{M \times N}, \tag{3.14}$$

$$\text{SD} = \sqrt{\frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(RT_{ij} - \text{Mean})^2}{M \times N}}, \tag{3.15}$$

$$\text{Skewness} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(RT_{ij} - \text{Mean})^3}{M \times N \times SD^3}, \tag{3.16}$$

$$\text{Kurtosis} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(RT_{ij} - \text{Mean})^4}{M \times N \times SD^4} - 3, \tag{3.17}$$

$$\text{Entropy} = \frac{\sum_{i=0}^{M}\sum_{j=0}^{N}(RT_{ij})^2}{M \times N}. \tag{3.18}$$

These 5 features are computed for each one of the 39 RTs obtained.

3.2.3 The VGG16 Deep Convolutional Neural Network

In this approach, the pre-trained deep CNN described in [38] was used. The chosen configuration was the D (Figure 3.9), with 16 layers - 13 convolution layers and 3 fully-connected layers - and the concept of Transfer Learning was applied.

The convolution layers are stacks of filters with small receptive field: 3×3 (smallest size to capture notions of left/right, up/down and center). The size of the receptive field of these filters differs from the ones of the top-performing entries of the ILSVRC-2012 and ILSVRC-2013 competitions where they used larger receptive fields, 11×11 and 7×7 respectively. The advantages of having stacked convolution layers is that makes the decision function more discriminative and by having a small receptive filter the number of parameters is small.

Between the 3 fully-connected layers, the first two layers have 4096 channels while the last one has 1000 channels, because it was the number of classes to be classified in ILSVRC challenge.

Transfer learning consists in transferring knowledge previously acquired in different domains than the one we are working on [39]. In the problematic of the work this means, that the CNN was trained in the database of ILSVRC-2014 challenge [38]

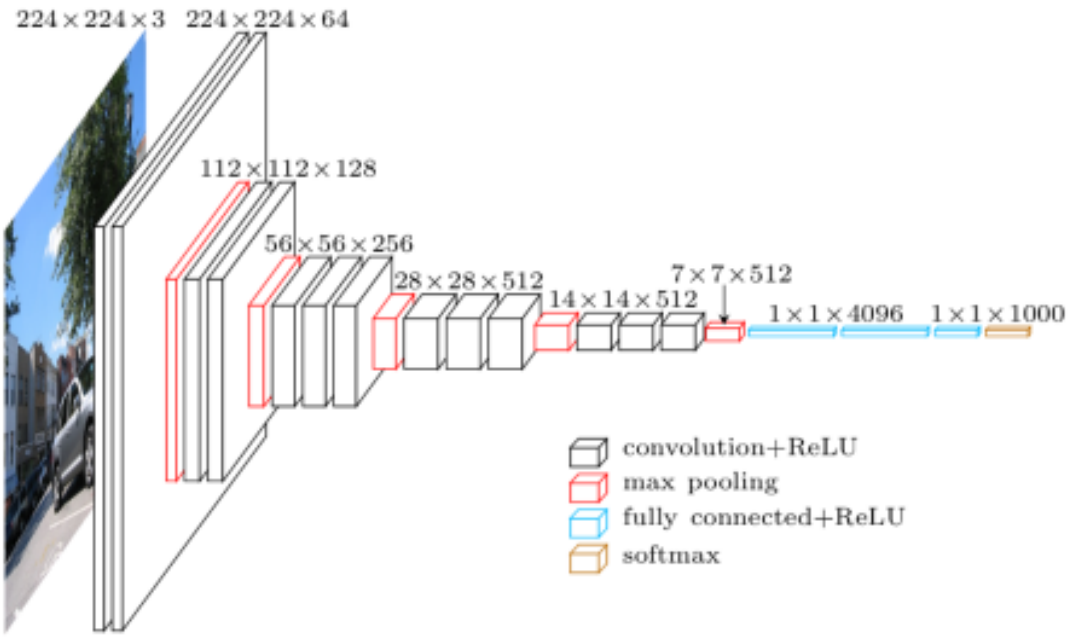


Figure 3.9: Architecture of VGG16 [http://www.cs.toronto.edu]

and was used as starting point for histopathological images. Instead of starting with random layer weights, the network training starts with the weights obtained from training on the ILSVRC-2014 challenge database. However, two modifications were made in the first and in the third layers of the FC layers. In the first layer, the training was started with random weights and in the third layer the number of channels was change to 2, the number of classes to be classified.

3.3 Stain Normalization and Color Deconvolution

In histopathology the diagnosis of a disease is based on visual examination of the tissues using a microscope. The increasing number of WSI scanners, requires the development of CAD systems to automatically analyze the WSIs, however these systems may face some complications due to the color variations in tissues. The color variations are the result of some factors like the brand of the scanners, chemical variations from different manufactures stains, staining procedure and section thickness [40].

In the set of extracted HPFs, color variations are visible and could interfere with classification. To mitigate color variations, it is possible to find in the literature

some stain normalization methods. Stain normalization principle is matching colors between two images, the source (Figure 3.10.a) and the reference (Figure 3.10.b). In this work, three color normalization methods were used. One of this methods is the Histogram Specification (or Matching) method [41], where from the source image I_{source}, and its corresponding histogram $p_x(x)$, a transformation $f(x)$ that gives a final image I_{final}, whose histogram is similar to the histogram of the reference image $p_z(z)$, is seek. Another method is the Reinhard's method [42], which consists at a first moment in a two-step conversion of the RGB signals to Ruderman's [43] perception-based color space $l\alpha\beta$, which minimizes correlation between channels for many natural scenes. First the RGB image is converted to LMS cone space

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3811 & 0.5783 & 0.0402 \\ 0.1967 & 0.7244 & 0.0782 \\ 0.0241 & 0.1288 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{3.19}$$

and then from LMS cone space to $l\alpha\beta$ color space,

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}. \tag{3.20}$$

Once the $l\alpha\beta$ space is obtained for both source and reference image, the mean and standard deviation of the three $l\alpha\beta$ channels are calculated. The mean values are subtracted from the data points,

$$\begin{aligned} l^* &= l - \langle l \rangle \\ \alpha^* &= \alpha - \langle \alpha \rangle \\ \beta^* &= \beta - \langle \beta \rangle \end{aligned} \tag{3.21}$$

Now data points are scaled using standard deviation values

$$\begin{aligned} l' &= \frac{\sigma_r^l}{\sigma_s^l} l^* \\ \alpha' &= \frac{\sigma_r^\alpha}{\sigma_s^\alpha} \alpha^* \\ \beta' &= \frac{\sigma_r^\beta}{\sigma_s^\beta} \beta^* \end{aligned} \tag{3.22}$$

Finally, using the following equations

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 \\ 0 & \frac{\sqrt{6}}{6} & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} \tag{3.23}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 4.4679 & -3.5873 & 0.1193 \\ -1.2186 & 2.3809 & -0.1624 \\ 0.0497 & -0.2439 & 1.2045 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}, \tag{3.24}$$

the $l\alpha\beta$ space is reconverted to RGB color space and the final normalized image is obtained. Macenko *et al.* [44] also developed a stain normalization method that consists, for each stain, in calculating the intensity histogram for all pixels that have a majority of that stain and the 99^{th} percentile of these intensity values are used as approximation of the maximum. Finally all the histograms are scaled to have the same pseudo-maximum and that way could be compared.

Following stain normalization, color deconvolution was also performed on all HPFs. Color deconvolution is vastly used in digital histopathology and consists in separating different stain components. In our case, a function provided by Khan *et al.* [40] was used to obtain two stain grey level images (i.e., hematoxylin only (Figure 3.10.j) and the eosin only (Figure 3.10.k) from the stain normalized hematoxylin-eosin (H&E) stained image.

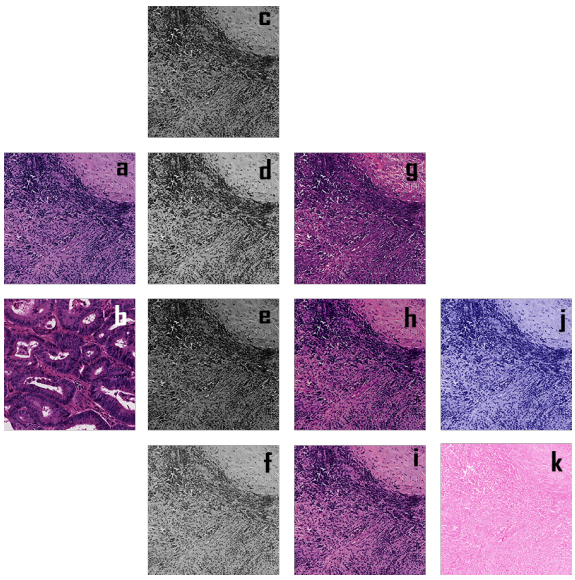


Figure 3.10: All the possible images obtained from the source image (a); b) reference image; c) intensity image; d-f) RGB channels respectively; g-i) Stain normalization methods (Histogram Specification, Reinhard's and Macenko's methods); j-k) Hematoxylin and Eosin stains

3.4 Support Vector Machine

Support Vector Machine (SVM) is a learning classification algorithm that learns from a training data set and attempt to generalize and make accurate predictions on new data sets. It is used for classification problems like binary classification [45]. In this work, SVM will be used to predict if a HPF is cancerous or non-cancerous.

The training data is composed by a set of input vectors, x_i, where each element of the vector corresponds to a $feature$ or $attribute$. Each vector is paired with a corresponding $label$, y_i, forming M pairs $(i = 1, ..., M)$ [45].

In the training step, usually the method aims to maximize the classification performance for the training data set. However if the classifier is too fit for the training data, the classification ability for novel data is negatively affected ($overfitting$) [2].

For simplicity, lets start by considering that data is linearly separable. That is, it is possible to split the training data in two classes by a hyperplane. In this way, the purpose of the SVM algorithm is to find the discrimination hyperplane. For M m-dimensional training inputs $x_i(i = 1, ..., M)$ where $y_i = 1$ for Class 1 and -1 for Class 2, the discriminant function is

$$D(x) = w^T x + b \tag{3.25}$$

where w is an m-dimensional vector and b is a bias term [2].

To control separability the following inequalities must be satisfied:

$$\begin{cases} w^T x_i + b \geq 1 & \text{for} \quad \text{I such that } y_i = 1, \\ w^T x_i + b \leq -1 & \text{for} \quad \text{I such that } y_i = -1. \end{cases} \tag{3.26}$$

which are equivalent to

$$y_i(w^T x_i + b) \geq 1 \quad \text{for} \quad i = 1, ..., M. \tag{3.27}$$

The hyperplane

$$D(x) = w^T x + b = c \quad \text{for} \quad -1 < c < 1 \tag{3.28}$$

forms a set of separating hyperplanes for each c that separates x_i $(i = 1, ..., M)$ by classes. The distance between the separating hyperplane and the training data that is closest to the hyperplane is designed by $margin$. The generalization region for the decision future is defined by $\{x|-1 \geq D(x) \geq 1\}$. The location of the separating hy-

perplane defines the generalization ability, being the *optimal separating hyperplane* the one with the maximum margin (Figure 3.11) [2].

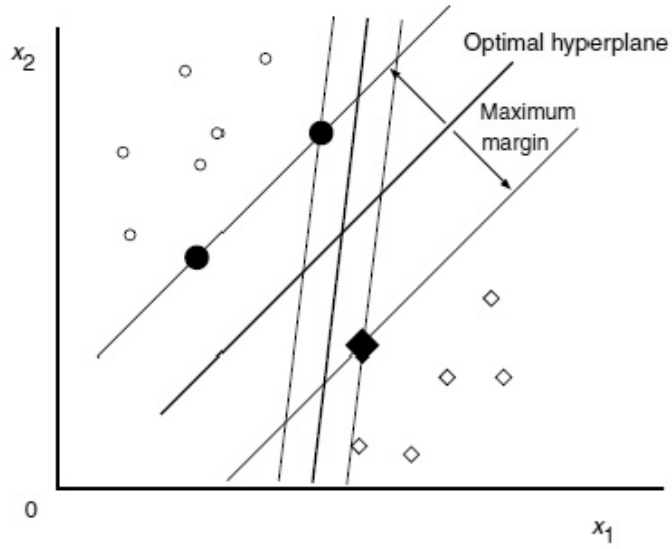


Figure 3.11: Optimal Separating Hyperplane [2].

3.5 K-Fold Cross-Validation

In order to estimate prediction error, cross-validation is the simplest and most widely used method. When the data available is not enough to divide it in two groups – training and validation – k-fold cross-validation is a good alternative [46].

It consists in divide the data in k equal groups, perform the train in (k-1) groups, the validation in the group that was left out and changing this group at each iteration. In the end the average value, of all k tests, is calculated [46].

3.6 Feature Selection

With the growth of the dimensionality and size of datasets machine learning methods face more problems when dealing with larger number of input features, being essential a pre-processing of the data, in order to find a projection of the data onto a smaller number of features, preserving the information as much as possible [47, 48]. Feature selection techniques are the answer because they detect the most relevant features, while removing irrelevant, redundant or noisy data [47]. Dimensionality

reduction is not the only objective of feature selection, they also try to avoid overfitting and provide faster cost-effective models [49]. In next subsections, some feature selection that exists in the literature will be presented and the one that was used in this work will be explained.

3.6.1 State of the Art

The feature selection techniques can be divided in three groups in relation to the way they work, each one with its advantages and disadvantages. These groups are classified as filter, wrapper and embedded [49, 50].

Filter methods look only for intrinsic properties of the data. Usually this type of methods ranks features by relevance score and remove those with a low score. The high-ranked features are presented to the classification algorithm after. As advantages they are easily scaled to high-dimensional datasets, they are computationally simple and fast and they are independent of the classification algorithm. The main disadvantages are the non interaction with the classifier (feature subset search separated form hypothesis search) and the majority of methods are univariate [49]. Methods like Chi-squared, Euclidean distance, information gain are included in univariate filter type. Correlation-based Feature Selection (CFS) [51], Markov Blanket Filter (MBF) [52] and Fast correlation-based feature Selection (FCFS) [53] are some of multivariate filter type methods.

Wrapper methods, contrary to filter methods, include the hypothesis search in the feature subset search and can be divided into two classes: deterministic and randomized. The subset of features is evaluated by training and testing a classification model. With the increasing of features' number, the optimal subset is obtained using heuristic search methods. The advantages of this type of method are the interaction between the feature subset search and the classifier and it takes in account the dependencies between features. They have two big drawbacks: higher probability of overfitting and they have a high computationally cost [49]. SFS [18, 54] Sequential Backward Elimination (SBE) [54], Beam search [55] and Genetic Algorithms [56] are amongst the wrapper type methods. In this work, the feature selection method used was SFS.

Finally, in the embedded methods type the optimal feature search is built in the classifier working as a search in the combined space of features subsets and hypotheses. They have as advantages being less expensive computationally speaking [49]. Weighted Naïve Bayes [57] is one of this type of methods.

3.6.2 Sequential Forward Selection

This algorithm starts with an empty subset of features, which is progressively filled with a feature not selected yet. The choice of this feature relies in a criterion value previously determined. The criterion choose in this work is the SVM classification accuracy obtained using a k-fold cross-validation (Section 3.5). In the end the average value, of all k tests, is calculated, in our case the classification accuracy. When this accuracy stops improving, the SFS algorithm completes and provides the best set of features.

3.6.3 Modified SFS

The modified SFS was used in the Camelyon16 challenge. The process is similar to that described in Section 3.6.2. The differences between the original and the modified version of SFS are the criterion, F-Score instead of accuracy and after a feature been added to the initial subset of best features, a correlation between the choose feature and the remaining features is calculated. Those that have a correlation with the selected features higher than 0.9 are eliminated from the process.

3.7 Evaluation Criteria

In a binary classification, some measures (Table 4.1) exist to understand how well a classifier works. These measures are calculated from the different elements of a confusion matrix – True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

A TP case is when our classifier classifies positively a positive case. Classifying positively a negative case is a FP. A positive case classified as negative is a FN. Finally, a TN is when the classifier classifies negatively a negative case.

3.8 Receiver Operating Characteristic (ROC) Curve

For the evaluation a Receiver Operating Characteristic (ROC) curve was used. A ROC curve is a two-dimensional graph in which is plotted the sensitivity (true

Table 3.1: Evaluation metrics

Measure	Formula	Evaluation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall effectiveness of the classifier
Precision	$\frac{TP}{TP+FP}$	Retrieved instances that are relevant
Recall (Sensitivity)	$\frac{TP}{TP+FN}$	Ability to identify positive instances
F-Score	$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision}+\text{recall}}$	Relation between positive cases and those given by the classifier
Specificity	$\frac{TN}{FP+TN}$	Ability to identify negative instances

positive rate) against 1-specificity (false positive rate) at different thresholds [58]. In this work, the threshold is the number of positive HPFs.

The main goal is to classify a LN as cancerous or non-cancerous based on the number of positive HPFs. Doing this way, it is possible to establish two different operation points. The practical one, that considers a LN cancerous if it has at least one positive HPF and the optimal one, that corresponds to the maximum number of positive HPFs that a LN can have to be considered non-cancerous. In both cases a F-score (performance measure for binary classification) is calculated relatively to the condition of the LN and the Area Under the Curve (AUC).

3.9 CAMELYON16

The challenge had two strategies to evaluate the algorithms.

The first one was the slide-based evaluation. Each participant should give at each slide the probability of containing metastasis. This was used to discriminate slides containing metastasis from normal slides. ROC analysis was performed and AUC was calculated to compare with other algorithms. The second one was the lesion-based evaluation. In this evaluation each participant should give X and Y coordinates of potential metastatic regions as well a confident score representing the probability of the detected region to be a tumor. A Free-Response Receiver Operating Characteristic (FROC) curve was used.

The strategy used, for the first evaluation to find the probability of a LN to contain metastases, was based in the threshold of the optimal operating point of our

database. Each time the number of cancerous classified HPFs exceed that threshold an average value of the probabilities of each cancerous glshpf was calculated.

For the second evaluation strategy, a special low-resolution binary image of the WSI was created. Each pixel in this low resolution image corresponds to a super pixel of size 800×800 (or HPF). At each pixel, a value of 1 (positive) or 0 (negative or background) is assigned depending on the classification. With a morphological operator any connected region larger or equal to 4 pixels were identified. Since each pixel corresponds to a HPF, the HPF with the highest probability was selected and the X and Y coordinates of its central point provided.

3.10 Main Pathways for Metastasis Detection in Histological Images

3.10.1 Diagrams

In Figure 3.12, the diagram representing the pathway of the method used is shown. The left side represents the DCNN while the right-side the traditional way. Briefly, in the traditional way the HPFs are extracted from the WSI annotations. For each HPF, nine images are derived and the GLCM and Laws features are extracted. The feature selection algorithm glssfs is applied to the features to obtain the optimal features, which in turn are used to validation of the method using k-fold cross-validation and to train the SVM-model. For the DCNN, the HPFs are extracted from the annotated WSIs and divided in two groups, one for train a DCNN and another to validate it.

In Figure 3.13 is the diagram for the testing part. The backbone of both approaches is the same, changing only the size of the HPFs (and consequently its number) and also the classifier. After classifying all the extracted HPFs, the cancerous ones are counted and the ROC curve obtained. From the ROC curve the operation points are obtained and the F-Score calculated for these points.

In Figure 3.14 the diagram shows the training path used for Camelyon16 challenge. It started with cancerous and non-cancerous WSIs. The HPF extraction follows the steps described in 3.1.2. From the ensemble of HPF extracted (around 55k) 2130 were selected randomly to validate and train a SVM-model. From here, the steps done are the same as for our database, with the exception of the feature selection

algorithm. The modified version was used.

The testing part of the Camelyon16 challenge is shown in Figure 3.15. The HPFs are extracted from the WSIs and the optimal features obtained in the training part are extracted from the HPFs obtained. After, the HPFs are classified based on the features extracted and the evaluations described in 3.9 performed.

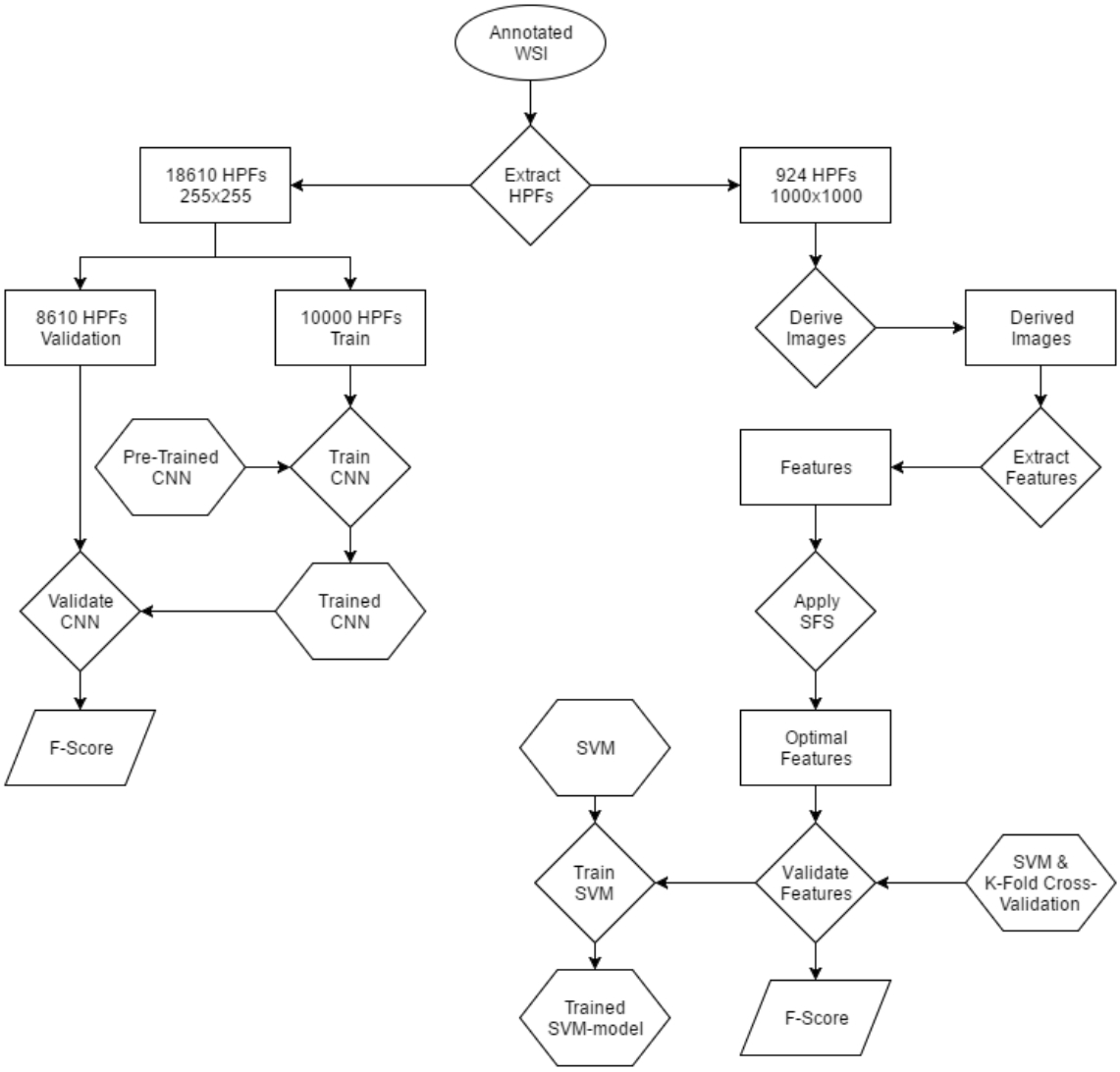


Figure 3.12: Diagram representing the training pathways of both methods (Traditional and DCNN).

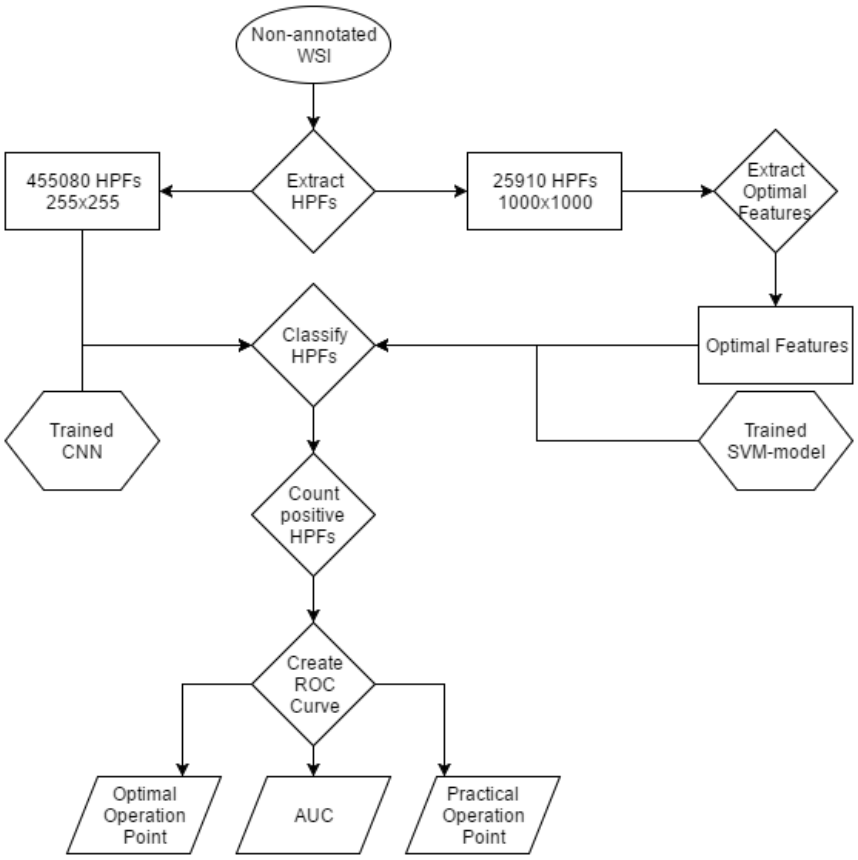


Figure 3.13: Test pathway. Right and left sides perform at different times.

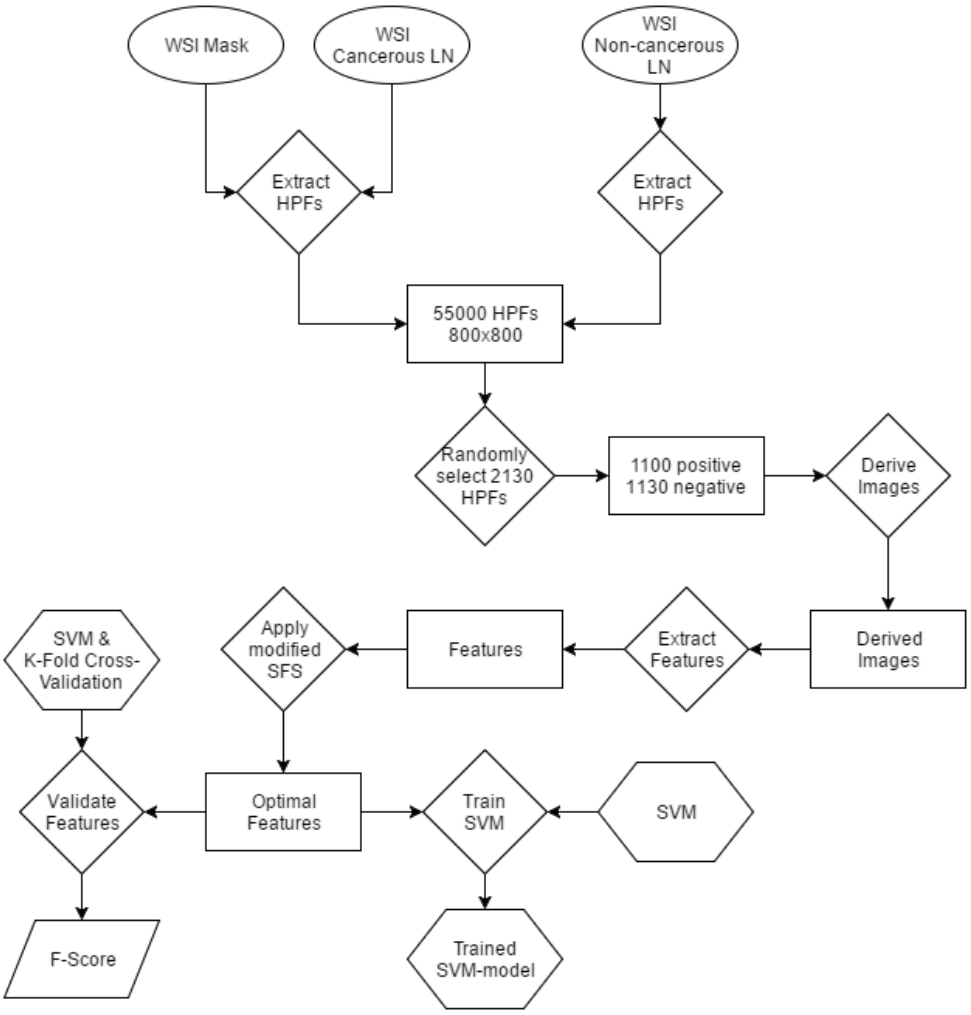


Figure 3.14: Camelyon16 challenge training pathway.

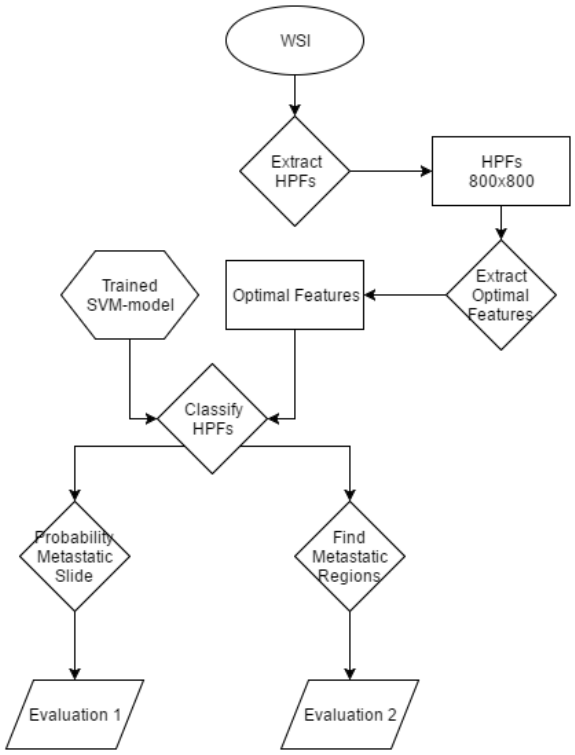


Figure 3.15: Evaluation path for Camelyon16 challenge.

3.10.2 Infrastructure

The configuration of the computer used for the DCNN calculations is a CPU Intel Xeon with 16 physical cores (32 in hyperthreading). A GPU Nvidia GeForce GTX TITAN X with 12Go of memory GDDR5 and 3072 cuda cores. Finally, it has 128Go DDR4 of memory RAM.

3.11 Correlation Between Histology and Ultrasounds

Our approach for the correlation had as starting point the results obtained with the more conventional texture analysis method. The final SVM-model used for test this method was based only in two features (Table 4) and were these two features that were used to correlate with the two QUS parameters presented earlier in Section 2.1. The two features will be known as L7L7_mean and L5L5_skew.

Like the QUS parameters it was necessary to obtain a value that described the whole LN. For each LN (annotated and non-annotated) we extracted the HPFs

as described in Section 3.1.1.3 and from each HPF we extracted the two chosen histologic parameters (both based on Laws energy texture measures). Once the features extracted the median value of all the HPFs was calculated for each LN. With 4 parameters for each LN (2 QUS and 2 Histological) we created a scatterplot matrix from where we proceeded with the analysis of a possible correlation between the two domains. The approach followed was to find a possible linear relation between the QUS and histology parameters. We started by normalizing the variables using the following formula

$$X_{norm} = \frac{X - \bar{X}}{\sigma_X} \tag{3.29}$$

where X is the set of values to be normalized, X is the mean value of all X and σ_X is the standard deviation of all X values. With the variables normalized, a scatter plot between the 2 QUS parameters and another between the 2 histological parameters were drawn. LDA boundaries were calculated and drawn in each scatter plot. To understand if the points that are close to the US boundary are also close to the Histology boundary another analysis was performed. A new scatter plot was created in which the axes are the QUS signed distance between the point and the QUS LDA boundary and the Histology signed distance between the point and the histological LDA boundary. Finally, 2 different situations were analyzed by performing the hypothesis test for regression slope by stating the following hypothesis

$$\begin{cases} H_0 : B_1 = 0 \\ H_a : B_1 \neq 0 \end{cases} \tag{3.30}$$

where B_1 is the slope of the regression line and the significance level value is 0.05. In the first situation we use both types of LNs together and in the second we do it separately.

4

Results

4.1 Detection of Metastatic Colorectal LN in Histological Images

4.1.1 Conventional Methods

Later, was said that the number of features (6435) that were extracted from all of the HPFs was quite large and to solve that issue, the feature selection algorithm SFS was used. Five features have come out from the algorithm, shown in the table. Using this five features and the training database (924 HPFs), all the possible combinations were tried between them to find which provided the best results using k-fold cross-validation ($k = 10$). The best result (F-score of 0.996 ± 0.005) was obtained using only two of the 6435 features. Both features were extracted from the Hematoxylin Stain (derived image), using Laws' $RT_{L_{\#}E_{\#}}$ (# equal to 5 and 7), from which the mean (size 7) and the skewness (size 5) were calculated.

Table 4.1: Optimal features obtained by using SFS

Derived Image	Mask applied to image	Statistical Descriptor	F-Score
Hematoxylin Stain	Laws' L7E7	Mean	0.996 ± 0.005
Hematoxylin Stain	Laws' L5E5	Skewness	
Intensity	Laws' S3S3	Kurtosis	
Intensity	Laws' L3E3	Entropy	
Intensity	Laws' L5E5	Standard Deviation	

Using the SVM model, trained using these two features, in the test dataset as explained above a ROC curve with an AUC of 0.98 was obtained (Figure 4.1).

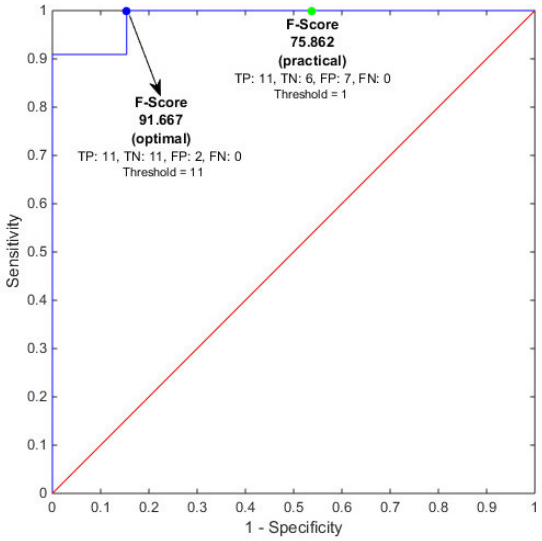


Figure 4.1: ROC curve for the traditional method.

4.1.2 Deep Convolutional Neural Networks

In contrast to the traditional approach, the validation of the classifier did not include a k-fold cross-validation. The DCNN was trained on a set of 10000 HPFs and validated on a set of 8610 HPFs (4955 positive and 3655 negative) with a F-score of 0.993.

Using the non-annotated database, the ROC curve of Figure 4.2 was obtained with an AUC of 1.0.

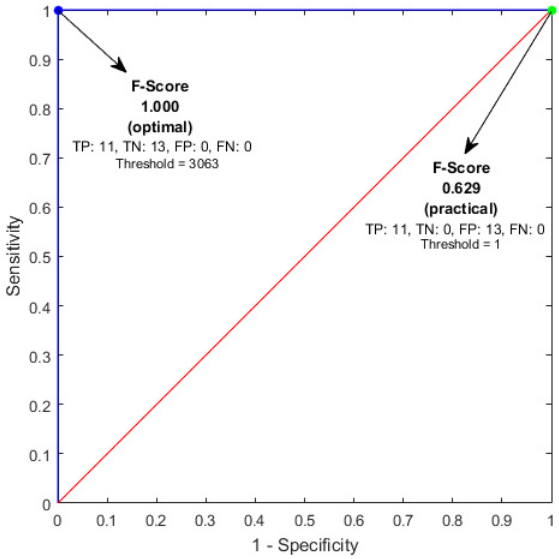


Figure 4.2: ROC curve of DCNN approach.

4.2 Correlation between Histology and Ultrasounds

The result of the scatterplot matrix is shown in Figure 4.3.

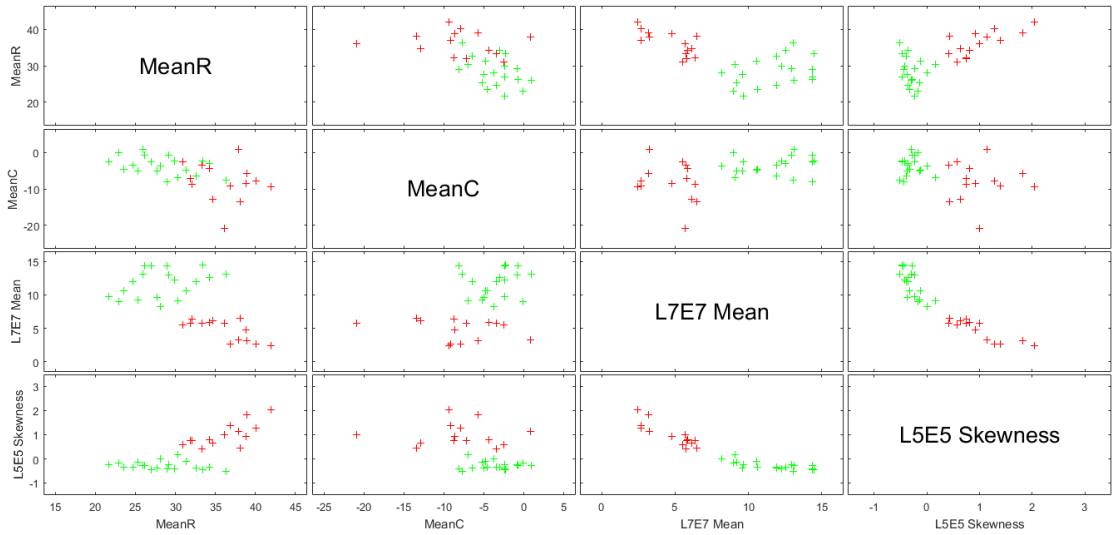
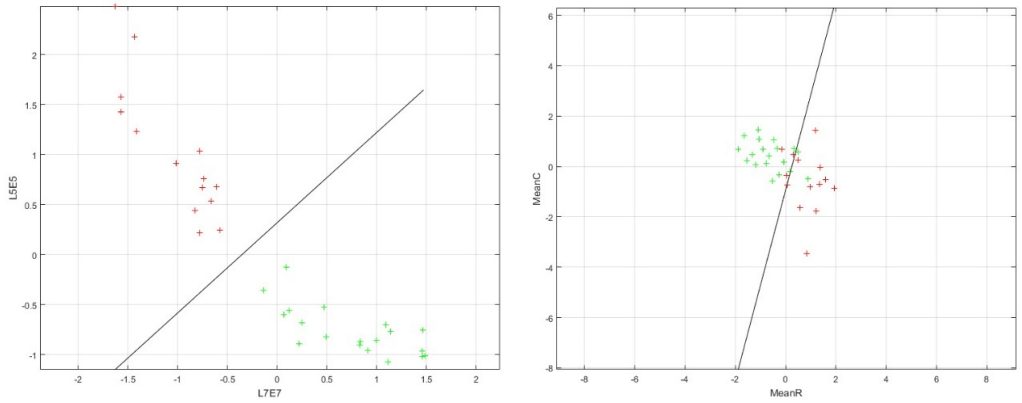


Figure 4.3: Scatterplot matrix

Using the scatter plots of positions (L2,C1) and (L4,C3) of Figure 4.3, a LDA was performed and the boundaries drawn (Figure 4.4).



(a) **(b)**

Figure 4.4: Scatter plots with LDA boundaries. At left between the histologic parameters and at right between the QUS parameters.

The graph in Figure 4.5 shows in x-direction the signed distances between the LNs

and the US LDA boundary and in the y-direction the signed distances between the LNs and the Histology LDA boundary.

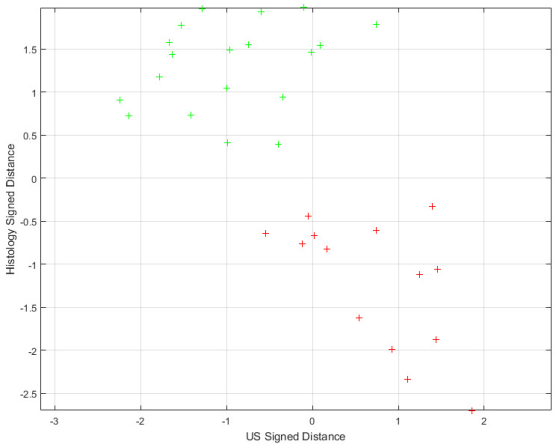


Figure 4.5: Scatter plot between the histologic signed distances and QUS signed distances

In Figure 4.6 is presented the linear regression for the first situation described in 3.11. The hypothesis test for the regression slope gave us a p-value of 0.00000476 with a R-squared of 0.496.

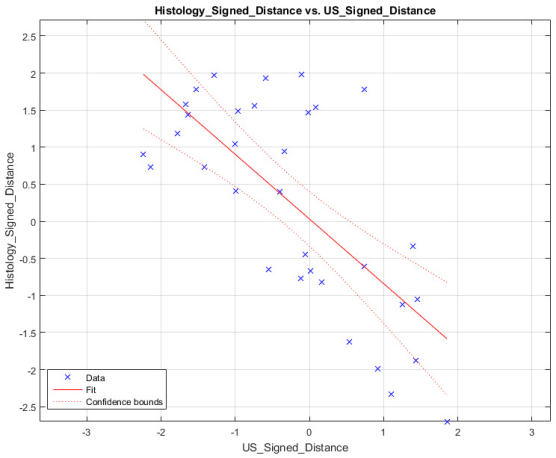
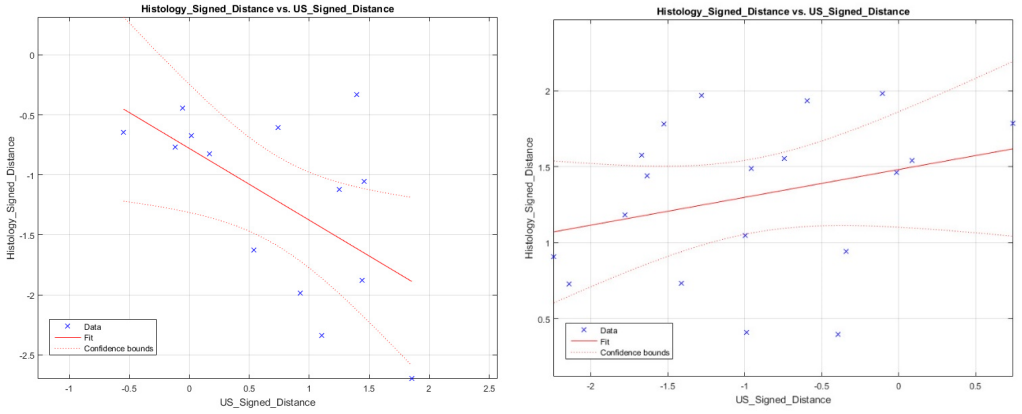


Figure 4.6: Linear Regression of both metastatic and non-metastatic LN

For the second situation we obtained a p-value of 0.0293 and a R-square of 0.338 for the metastatic LNs (Figure **??**) and for the non-metastatic LNs (Figure 4.7b) a p-value of 0.227 and a R-squared of 0.085.



(a) **(b)**

Figure 4.7: Linear Regression of metastatic LNs and non-metastatic LNs.

4.3 Detection of Metastatic Breast Lymph Node on Histological Images (CAMELYON16)

The results of the training and validation using 2130 HPFs. The modified SFS algorithm that we use in the challenge gave us an output of 8 features, presented in Table 4.2.

Table 4.2: Modified SFS algorithm output (CAMELYON16). In last column, O stands for orientation angle, D for interpixel distance and K for kernel used.

Derived Image	Feature Base	Feature	Parameters
Hematoxylin Stain	GLCM	Inverse Difference Moment	$O : 135°$ $D : 9$
Hematoxylin Stain	GLCM	Inverse Difference Moment	$O : 135°$ $D : 9$
Reinhard Color Normalization	GLCM	Entropy	$O : 90°$ $D : 10$
Blue Channel	GLCM	Difference Variance	$O : 135°$ $D : 2$
Eosin Stain	GLCM	Correlation	$O : 90°$ $D : 7$
Reinhard Color Normalization	GLCM	Sum Entropy	$O : 90°$ $D : 8$
Red Channel	Laws	Entropy	$K : O7O7$
RGBHist Color Normalization	GLCM	Sum Entropy	$O : 90°$ $D : 9$
Intensity	GLCM	Inverse Difference Moment	$O : 90°$ $D : 10$

Like we did for the colorectal database, we tested all the feature combinations between the 8, using a k-fold cross-validation. Using the 8, it was the combination that produced the best result, with an F-Score of 0.838 ± 0.024. We do not have access to the ground truth of the test database, however the organization made the results public in their website. A total of 24 research groups have participated in the challenge. For the slide-base evaluation we came last with a AUC of 0.5636. For the lesion-based evaluation, we came second-to-last with an AUC of 0.1152. The big winner of the challenge was the Harvard Medical School and Massachusetts Institute of Technology, USA, they obtained the first place in both evaluations with AUC of 0.9250 and 0.7051.

5

Discussion and Conclusions

In this work we developed two different approaches to detect metastatic regions in histology and we correlate some parameters of one of the approaches with a couple of QUS parameters. The two approaches used in the analysis and classification of the histological images were a CAD system based on conventional texture methods while the other one uses deep convolutional neural networks.

In terms of performance of the histological approaches, the principal criterion to choose the operation points of this systems were guaranteeing the existence of zero false negative cases. The practical one is the less flexible because in the presence of one positive HPF it considers the LN as cancerous. The optimal one, gives to the classifier some flexibility because it is based on a larger threshold of positive classified HPFs.

At the practical operation point, the best system, based on F-Score, is the texture-based one (0.759 vs 0.629). Using it, one may exempt 6 of the 24 LNs from pathologist examination because they do not have any positive HPFs. The DCNN-based system has a poor performance at this point, due to the fact that all LNs have positive classified HPFs. A possible explanation to this fact could be the size of the HPFs and consequently the number, one can extract from each LN, be higher in the DCNN-based system. Since the classifier is not perfect, the probability to make a mistake on at least one HPF increases with the number of HPFs.

Talking in terms of the optimal operation point, the best system is the DCNN-based system (F-Score of 1 against 0.917 of texture-based one). The threshold of this point is very different in both cases. Eleven with the texture-based and 3063 with the DCNN-based classifier. Like in the case of the other operation point the number of extracted HPFs has influence in the threshold, but it could not be the only reason. Not all the LNs have the same number of WSIs and consequently the same number of sections were the HPFs can be extracted, which ranges from 3 to 20. Adding the missing sections to the LNs, balancing the number between them

as well as adding more LNs to the database would improve the robustness of the threshold. For the DCNN-based system the size of the HPFs was smaller because this type of networks needs a large number of images. The disadvantage of this size of HPF is that the field of view is more limited. A possible way of solving it is increasing the number of HPFs and enlarging them then creating new HPFs by rotating and slightly deforming existing WSIs.

In Section 3.10 are described all the steps for both of the approaches. Even if the time spent was not recorded, the DCNN-based training approach was largely faster than the traditional approach – around one day against 7 days (between 5/6 days to feature selection).

In terms of correlation between the QUS parameters and the histological parameters our goal was to find a linear correlation between them to allow the progression towards the objective of guide the pathologist to a specific region of the LN before cutting it. So we studied the linear regressions between the signed distances (distances from the points in the graphs of Figure 4.4 to the respective LDA boundary) of QUS parameters against histological parameters. We did two different analyses. One for both types of LNs (Figure 4.6) and one for metastatic LNs (Figure 4.7a) and non-metastatic LNs (Figure 4.7b). By analyzing both LN types together and basing our judgment only on the p-value, we can reject the null hypothesis ($0.00000476 > 0.05$). However, when we analyzed the cases (metastatic and non-metastatic) separately, a completely different story happened. Because we did two tests, we divided the significance level by two. Both p-values of metastatic and non-metastatic LNs were 0.0295 and 0.227 which are bigger than 0.025 (significance level). We could not reject the null hypothesis so we conclude that there is not a linear relation between QUS and histological signed distances.

We thought in another way of getting a relation/correlation between the two data sources. It consisted on using the graphs of the positions (L4,C1) and (L3,C4) of Figure 4.3, find a polynomial regression between $MeanR$ and L5L5_skew variables, a polynomial regression between the two histological parameters and use the LDA discriminant function of Figure 4.4a. Using the value of $MeanR$, directly extracted from the US we could predict a value for L5L5_skew, which in turn it would be used to predict a value for L7L7_mean. Finally, using the LDA discriminant function it would be possible to predict the state of the LN.

The two attempts of finding a correlation between the QUS and Histologic parameters failed. In future attempts, another type of approaches could be tried. For example, we used the same four parameters all time and we focused our efforts in

the classification only, being necessary to change the approach, the parameters and even using other types of texture analysis methods. It would be also important to increase the database.

The results obtained in the CAMELYON16 challenge were not the expected, since we obtained very good results in our database. As possible causes we have the type of cancer and the resolution of the images. In the challenge they have used sentinel lymph nodes of patients with breast cancer to obtain the images. These images were digitized at 40× (20× in colorectal database) and the HPFs were extracted from the 40× images.

In conclusion, it is possible to use texture to locate micrometastases in LNs and that way guide the pathologist to a specific region in the WSI (Figure 5.1). The results obtained in both the approaches are highly satisfactory. It is impressive that old texture methods give very good results, even better than those obtained using DCNNs – considered the future of machine learning for classification problems. We could not find a correlation between the QUS and histologic parameters. Or that correlation does not exist or the approaches were not the best. The CAMELYON16 challenge would take place again the next year so it could be nice gather a team, solve the problems of this edition and try again.

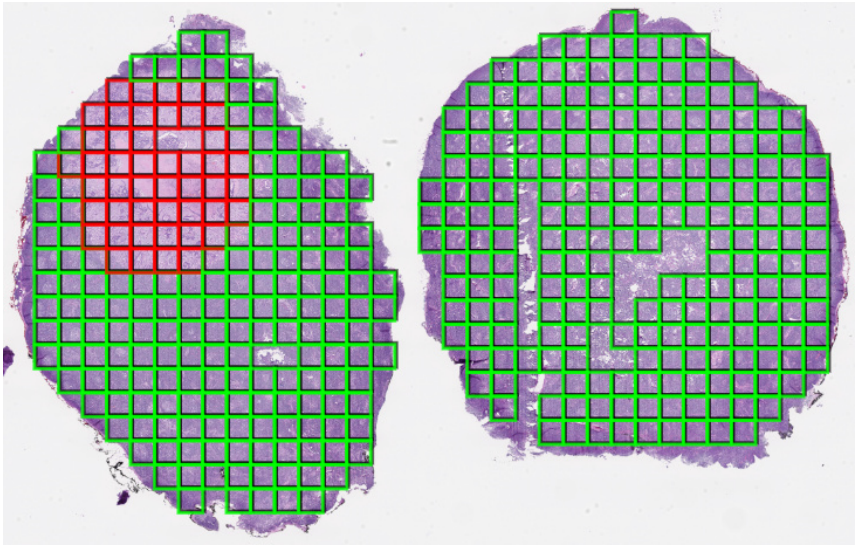


Figure 5.1: Result of HPF classification. A metastatic region (red) is clearly visible and well separated from the non-metastatic regions (green)

References

[1] M. H. Ross and W. Pawlina, *Histology : a text and atlas : with correlated cell and molecular biology*. 2006.

[2] S. Abe, *Support Vector Machines for Pattern Classification (Advances in Pattern Recognition)*, vol. 224. Springer-Verlag New York, Inc., 2005.

[3] National Cancer Institute, "What is Cancer?," accessed May 16, 2016. `http://www.cancer.gov/about-cancer/what-is-cancer`.

[4] American Cancer Society, "Lymph Nodes and Cancer."

[5] E. B. Rankin, J. Erler, and A. J. Giaccia, "The Cellular Microenvironment and Metastases," in *Abeloff's Clinical Oncology*, pp. 40–51.e5.

[6] T. Ookubo, Y. Inoue, D. Kim, H. Ohsaki, Y. Mashiko, M. Kusakabe, and M. Sekino, "Characteristics of magnetic probes for identifying sentinel lymph nodes," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2013, pp. 5485–5488, 2013.

[7] M. L. H. Ong and J. B. Schofield, "Assessment of lymph node involvement in colorectal cancer," *World Journal of Gastrointestinal Surgery*, vol. 8, no. 3, pp. 179–192, 2016.

[8] C. Verrill, N. J. Carr, E. Wilkinson-Smith, and E. H. Seel, "Histopathological assessment of lymph nodes in colorectal carcinoma: does triple levelling detect significantly more metastases?," *Journal of clinical pathology*, vol. 57, no. 11, pp. 1165–7, 2004.

[9] E. Saegusa-Beecroft, J. Machi, J. Mamou, M. Hata, A. Coron, E. T. Yanagihara, T. Yamaguchi, M. L. Oelze, P. Laugier, and E. J. Feleppa, "Three-

dimensional quantitative ultrasound for detecting lymph node metastases," *Journal of Surgical Research*, vol. 183, no. 1, pp. 258–269, 2013.

[10] J. Mamou, A. Coron, M. L. Oelze, E. Saegusa-Beecroft, M. Hata, P. Lee, J. Machi, E. Yanagihara, P. Laugier, and E. J. Feleppa, "Three-dimensional high-frequency backscatter and envelope quantification of cancerous human lymph nodes.," *Ultrasound in medicine & biology*, vol. 37, no. 3, pp. 345–57, 2011.

[11] J. Mamou, A. Coron, M. Hata, J. Machi, E. Yanagihara, P. Laugier, and E. J. Feleppa, "Three-dimensional High-frequency Characterization of Cancerous Lymph Nodes," *Ultrasound in Medicine & Biology*, vol. 36, no. 3, pp. 361–375, 2010.

[12] A. Goode, B. Gilbert, J. Harkes, D. Jukic, and M. Satyanarayanan, "OpenSlide: A vendor-neutral software foundation for digital pathology," *Journal of pathology informatics*, vol. 4, no. 1, 2013.

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[14] D. Liu and J. Yu, "Otsu method and K-means," *Proceedings - 2009 9th International Conference on Hybrid Intelligent Systems, HIS 2009*, vol. 1, no. 2, pp. 344–349, 2009.

[15] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, "Multi-class texture analysis in colorectal cancer histology," *Scientific Reports*, vol. 6, p. 27988, 2016.

[16] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1205–1218, 2012.

[17] Y. Freund and R. R. E. Schapire, "Experiments with a New Boosting Algorithm," *International Conference on Machine Learning*, pp. 148–156, 1996.

[18] S. Bouatmane, M. A. Roula, A. Bouridane, and S. Al-Maadeed, "Round-Robin sequential forward selection algorithm for prostate cancer classification and diagnosis using multispectral imagery," *Machine Vision and Applications*, vol. 22, no. 5, pp. 865–878, 2010.

[19] K. Nguyen, B. Sabata, and A. Jain, "Prostate cancer detection: Fusion of cytological and textural features," *Journal of Pathology Informatics*, vol. 2, p. 3, 2011.

[20] H. A. Elnemr, "Statistical Analysis of Law ' s Mask Texture Features for Cancer and Water Lung Detection," *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 6, pp. 196–202, 2013.

[21] M. Rachidi, A. Marchadier, C. Gadois, E. Lespessailles, C. Chappard, and C. L. Benhamou, "Laws' masks descriptors applied to bone texture analysis: an innovative and discriminant tool in osteoporosis.," *Skeletal radiology*, vol. 37, pp. 541–548, 2008.

[22] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Computer Vision and Pattern Recognition*, 2015.

[25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[26] K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. J. Snead, and N. M. Rajpoot, "Gland Segmentation in Colon Histology Images: The GlaS Challenge Contest," *CoRR*, vol. abs/1603.0, pp. 1–24, 2016.

[27] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for Nuclei Detection, Segmentation and Classification in Digital Histopathology: A Review. Current Status and Future Potential," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 97–114, 2014.

[28] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8150 LNCS, no. PART 2, pp. 411–418, 2013.

[29] H. Su, F. Xing, X. Kong, Y. Xie, S. Zhang, and L. Yang, "Robust Cell Detection and Segmentation in Histopathological Images Uusing Sparse Reconstruction and Stacked Denoising Autoencoders," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 383–390, 2015.

[30] K. Sirinukunwattana, S. Raza, Y.-w. Tsang, D. Snead, I. Cree, and N. Rajpoot, "Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1196–1206, 2016.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015.

[32] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation," 2016.

[33] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast Cancer Histopathological Image Classification using Convolutional Neural Networks," *International Joint Conference on Neural Networks (IJCNN 2016)*, 2016.

[34] G. Litjens, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, P. Bult, B. V. Ginneken, and J. V. D. Laak, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Nature Publishing Group*, 2016.

[35] B. Nielsen, F. Albregtsen, and H. E. Danielsen, "Statistical nuclear texture analysis in cancer research: a review of methods and applications," *Crit Rev Oncog*, vol. 14, no. 2-3, pp. 89–164, 2008.

[36] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[37] K. I. Laws, *Textured Image Segmentation*. PhD thesis, University of Southern California, 1980.

[38] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ImageNet Challenge*, pp. 1–10, 2014.

[39] M. Oquab, M. Oquab, I. Laptev, J. S. Learning, T. Mid-level, M. Oquab, and L. Bottou, "Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks," *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1717–1724, 2014.

[40] A. M. Khan, N. Rajpoot, D. Treanor, and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1729–1738, 2014.

[41] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing*. 2010.

[42] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.

[43] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," *Journal of the Optical Society of America A*, vol. 15, no. 8, p. 2036, 1998.

[44] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," *Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009*, pp. 1107–1110, 2009.

[45] C. Campbell and Y. Ying, *Learning with Support Vector Machines*, vol. 5. 2011.

[46] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," *Elements*, vol. 1, pp. 337–387, 2009.

[47] V. Kumar and S. Minz, "Feature Selection: A literature Review," *Smart Computing Review*, vol. 4, no. 3, pp. 211–229, 2014.

[48] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Advances in Bioinformatics*, vol. 2015, no. 1, 2015.

[49] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[50] A. Jovic, K. Brkic, and N. Bogunovic, "A review of feature selection methods with applications," *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on*, no. May, pp. 1200–1205, 2015.

[51] M. a. Hall, "Correlation-based Feature Selection for Machine Learning," *Methodology*, vol. 21i195-i20, no. April, pp. 1–5, 1999.

[52] D. Koller and M. Sahami, "Toward optimal feature selection," pp. 284–292, 1996.

[53] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy," *Journal of Machine Learning Research*, vol. 5, pp. 1205–1224, 2004.

[54] J. Li, M. T. Manry, P. L. Narasimha, and C. Yu, "Feature selection using a piecewise linear network.," *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 17, no. 5, pp. 1101–15, 2006.

[55] W. Siedlecki and J. Sklansky, "On automatic feature selection," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 02, no. 02, pp. 197–220, 1988.

[56] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.

[57] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.

[58] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.