UNIVERSIDAD NACIONAL DE COLOMBIA

QUANTIFICATION OF GLANDS IN GASTRIC CANCER - CUANTIFICACIÓN DE GLÁNDULAS EN IMÁGENES HISTOPATOLÓGICAS DE CÁNCER GÁSTRICO

Sunny Catalina Alfonso Niño

Universidad Nacional de Colombia
Medical School, Department of Diagnostic Images
Bogotá D.C., Colombia
2019

QUANTIFICATION OF GLANDS IN GASTRIC CANCER - CUANTIFICACIÓN DE GLÁNDULAS EN IMÁGENES HISTOPATOLÓGICAS DE CÁNCER GÁSTRICO

Sunny Catalina Alfonso Niño

Thesis presented as partial requirement for the degree of:
Master in Biomedical Engineering

Advisor:
MD, MSC, PHD, Doctor, Medical Surgeon, Magister in Electrical Engineering, PhD in Sciences Biomedicales, Eduardo Romero Castro

Research Area:
Processing signals
Research Group:
Cim@lab

Universidad Nacional de Colombia
Medical School, Department of Diagnostic Images
Bogotá D.C., Colombia

2019

Choose a job you love, and you will never have to work a day in your life
Confucio

Life is like riding a bicycle. In order to keep your balance, you must keep moving
Albert Einstein

A goal is a dream with a deadline
Napoleon Hill

Acknowledgment

Mainly to Professor Eduardo Romero for his dedication and constant motivation in his desire to transform the reality of the country with education for all the knowledge I received every day, for his patience, effort, dedication, and passion in the development of his work and, especially, for their valuable friendship. To Ricardo Moncayo Mg. and Germán Corredor PhD. for his constant advice, for transmitting his knowledge and for immense love he has for this research topic, which allowed to consolidate this great work. I thank doctor and pathologist Angel Yobany Sanchez for many hours of advice transmitting their knowledge. I thank Ana Maria Morales who provided the samples and information needed about patients to do this research. I thank all the members of the Cimalab research group of the Universidad Nacional de Colombia who contributed their ideas for the development of this research. Finally, a special thanks to my friends who accompanied me in this process and to the most important people in my life, my family.

Content

List of Symbols, Images and Tables

Abbreviations

Abbreviation	**Terminus**
ANN	Artificial Neural Network
cm	Centimeters
DL	Deep Learning
DNA	Deoxyribonucleic Acid
FoV	Fields of View
GC	Gastric Cancer
GBT	Gradient Boosting Tree
$GBRT$	Gradient-Boosted-Regression-Tree
$H.pylori$	Helicobacter Pylori
$H\&E$	Hematoxylin and Eosin
$ILFS$	Infinite Latent Feature Selection
$mRMR$	Minimum Redundance Maximum Relevance
$NLCI$	Nuclear Local and Contextual Information
NS	No survival
$OLGA$	Operative Link on Gastritis Assessment
QDA	Linear & Quadratic Discriminant Analysis
ROC	Receiver Operating Characteristics
$R-CNN$	Region-based Convolutional Neural Network
RoI	Region of Interest
RNN	Recurrent Neural Network
S	Survival
$SPIE$	International Society for Optics and Photonics
SRN	Survival Recurrent Network
SVM	Support Vector Machine
USD	United States Dollars
WHO	World Health Organization
WSI	Whole Slide Images

List of Images

List of Tables

Resumen

La detección y cuantificación automática de las glándulas en el cáncer gástrico puede contribuir a medir objetivamente la gravedad de la lesión, desarrollar estrategias para el diagnóstico precoz y lo que es más importante, mejorar la clasificación del paciente; sin embargo, su cuantificación es una tarea altamente subjetiva, propensa a errores debido al alto tráfico de biopsias y a la experiencia de cada experto. La presente disertación de maestría está compuesta por tres capítulos los cuales llevan a la cuantificación objetiva de glándulas. En el primer capítulo del documento se presenta un nuevo enfoque para la segmentación de los núcleos glándulares en base a la información nuclear local y contextual (vecindario) "NLCI". Se entreno un Gradient-Boosted-Regression-Tree para distinguir entre núcleos glándulares y núcleos no glándulares. La validación se llevó con 45.702 núcleos anotados manualmente de 90 campos de visión (parches) extraídos de imágenes de biopsias completas de pacientes diagnosticados con cáncer gástrico. Finalmente, un modelo Deep Learning fue entrenado como línea base para comparar nuestros resultados. NLCI logró una precisión de 0.977 % y un F-Score de 0.955 %, mientras que la red convolucional "fast R-CNN" arrojó una precisión de 0.923 % y un F-Score y 0.719 %. En el segundo capítulo se presenta un marco completo para la detección automática de glándulas en imágenes de cáncer gástrico. Las glándulas candidatas de una versión binarizada del canal de hematoxilina, luego, la forma y los núcleos de las glándulas se caracterizan mediante características locales que alimentan un clasificador Random-Cross-Validation, entrenado previamente con imágenes anotadas manualmente por un experto. La validación se realizó mediante un conjunto de datos con 1.330 parches extraídos de siete biopsias de pacientes diagnosticados con cáncer gástrico. Los resultados mostraron una precisión del 93 % utilizando un clasificador lineal. Finalmente, en el tercer capítulo analiza las características más relevantes entre las glándulas y sus núcleos glandulares para predecir la sobrevida a un año de un paciente diagnosticado con cáncer gástrico. Una selección de características basada en información mutua: criterios de dependencia máxima, máxima relevancia y mínima redundancia (mRMR) escogen las características correlacionadas con la supervivencia del paciente. Se extrajo un conjunto de datos con 668 campos de visión (FoV), 2.076 estructuras glandulares de 14 imágenes completas de pacientes diagnosticados con cáncer gástrico. Los resultados mostraron una precisión del 78.57 % usando un Análisis Discriminante Lineal y Cuadrático (QDA) y un esquema de evaluación entrenando con trece casos y dejando un caso aparte para validar.

Palabras clave: Glándulas, cáncer gástrico, segmentación de núcleos, información local y contextual, mRMR, predecir, cuantificación, fast R-CNN, supervivencia.

Abstract

Automatic detection and quantification of glands in gastric cancer may contribute to objectively measure the lesion severity, to develop strategies for early diagnosis, and most importantly to improve the patient categorization; however, gland quantification is a highly subjective task, prone to error due to the high biopsy traffic and the experience of each expert. The present master's dissertation is composed by three chapters that carry to an objective identification of glands. In the first chapter of this document we present a new approach for segmentation of glandular nuclei based on nuclear local and contextual (neighborhood) information "NLCI". A Gradient-Boosted-Regression-Tree classifier is trained to distinguish between glandular nuclei and non glandular nuclei. Validation was carried out using 45.702 annotated nuclei from 90 fields of view (patches) extracted from whole slide images of patients diagnosed with gastric cancer. NLCI achieved an accuracy of 0.977 and an F-measure of 0.955, while R-CNN yielded corresponding accuracy and F-measures of 0.923 and 0.719, respectively. In second chapter we presents an entire framework for automatic detection of glands in gastric cancer images. By selecting gland candidates from a binarized version of the hematoxylin channel. Next, the gland's shape and nuclei are characterized using local features which feed a Random-Cross-validation method classifier trained previously with images manually annotated by an expert. Validation was carried out using a data-set with 1.330 from seven fields of view extracted from patients diagnosed with gastric cancer whole slide images. Results showed an accuracy of 93 % using a linear classifier. Finally, in the third chapter analyzing gland and their glandular nuclei most relevant features, since predict if a patient will survive more than a year after being diagnosed with gastric cancer. A feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy "mRMR" approach selects those features that correlated better with patient survival. A data set with 668 Fields of View (FoV), 2.076 glandular structures from 14 whole slide images were extracted from patient diagnosed with gastric cancer. Results showed an accuracy of 78.57 % using a QDA Linear & Quadratic Discriminant Analysis was training with Leave-one-out e.g training with thirteen cases and leaving a separate case to validate.

Keywords: Glands, gastric cancer, nucleus segmentation, local and contextual information, mRMR, predict, quantification, fast R-CNN, survival.

1 *Introduction*

Gastric cancer (GC) continues being a major public health problem. Especially considering that there are about one million new cases have been reported yearly worldwide. Which makes it the fourth most common cancer and the seven leading cause of cancer deaths around the world [8]. This disease is usually caused by genetic conditions, environmental risks, geographical conditions, bacterial infections, obesity, smoking and dietary habits. In general, GC is more common in older adults, having an incidence of 17 to 48 cases per 100,000 inhabitants. Consequently, the increase in mortality indicators corroborates that life expectancy at 5 years after the cancer is detected, is only of 10 % [41].
Etymologically, GC has been linked to infection with Helicobacter Pylori virus (HPV) [58] in spite of this fact, incidence has decreased worldwide due to prevention strategies and early diagnosis. In our country, more than 90 % of these cases have a late diagnosis. Incidence mortality in populations with HPV, begins with an allergy that ends up being a gastritis in 40 % of population, of which 10 % develops ulcer by preventing the correct absorption of nutrients, and finally 5 % of this population incurs suffering GC. Children under 10 years old is the population that is most likely to suffer this kind of infections, because the risks could be increase when they swim in dirty pools or rivers, as well as ingesting water or food not hygienically prepared.
Epithelial cells that overlay the gland usually have a white color and are affected by polluting substances contained in alcohol, different toxins in food and poor nutrition. In all case this will be affect the chromosomes of this cell, causing carcinogenic processes e.g. enterocytes, nitrosamines, aimes, nitrosamines, that can act on genetic material. These bad habits lead to the blockage of the regulatory mechanisms that control the cell cycle, and do not allow cells under normal conditions to have a controlled reproduction process. However, immune system of human being generates new renewing cells that cure possible infections in mucosa, every 15 to 21 days; for this reason it is possible that next state of cancer will not be continued, but that state of health will improve before metaplasia occurs.

1.1. Gastric Cancer States

Gastric cancer remains one of the deadly diseases with poor prognosis. New classification of gastric cancers based on histologic features, genotypes and molecular phenotype helps better understand the characteristics of each sub-type, and improve early diagnosis, prevention and treatment [40]. According to its extension or depth, GC is diagnosed dividing into early and advanced cancer; When a biopsy is diagnosed as early cancer it is understood that the tumor is within the mucosa and even inside of the submucosa and has a diameter minor than 5 cm; or, conversely, advanced GC already begins to exceed the submucosa measuring more than 5 cm of diameter and suggest ganglionic compromise mutating to other tissues of the body. The diagnosis was made by pathologists, who through the slide can analyze visual changes suffered by glands through the different states of GC [3]. Figure **1-1** shows stages of cancer evolution.

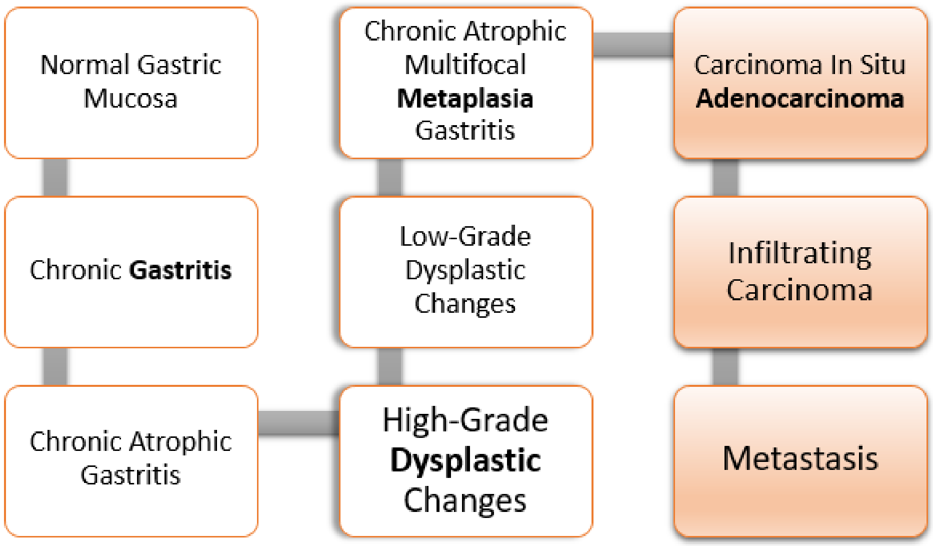

Figure 1-1: Evolution of GC, authorship.

Stomach is composed of glands and foveolas in which metaplasia, dysplasia and cancer can occur; within GC classes, the most advanced stage is known as adenocarcinoma. According to the Laurén classification, gastric adenocarcinomas are divided into intestinal, diffuse, mixed and indeterminate subtypes [45]. They vary not only in morphology but also in epidemiology, progression pattern, genetics and clinical picture. The main reason that this dissertation has, is research the tubular subtype one, occurs in about 54 % of the cases, as it is the most frequent to develop gastric cancer and it manifests the degrees of severity in terms of geometric characteristic in each cancer state. It is twice as often in males as in females and is localized mostly in the antrum. Histopathologically, it is characterized by malignant epithelial cells that show cohesiveness and glandular differentiation infiltrating the surrounding tissue [48]. Specifically in adenocarcinoma stage, the tubular structures are not objectively quantified to achieve a good differentiation or diagnosis according with the WHO "World Health Organization" this is because the quantification of glands might help to determine the degree of cancer; however, as it mainly relies on the visual interpretation of a pathologist, it may lead to a certain level of subjectivity. That is why automated tools could help to standardize diagnostic processes [10][47].

By contrast, the diffuse subtype (32 %) is characterized by tumor cells that show poor differentiation and lack of cohesion. This subtype occurs equally often in males and females and these patients are on average younger than those with intestinal GC. Intestinal type of gastric cancer is felt to be caused mainly by environmental (exogenous) factors whereas the diffuse type is thought to be due to hereditary and genetic (endogenous) factors. The intestinal and diffuse GC subtypes are pathologically considered as separate entities, but clinically, both are treated similarly. The main clinical difference is related to the different recurrence patterns, with the diffuse-mixed types more prone to peritoneal dissemination, especially when the serosa is involved, whereas the risk of liver metastases is higher in the intestinal type[10].

1.2. WHO classification

Compared to the Laurén´s system, the WHO classification is based on pure histo-morphological appearance. The WHO divides GC into tubular, papillary, mucinous, poorly cohesive (including signet ring cell carcinoma) and mixed carcinomas. This classification includes, besides adenocarcinomas, also all other types of gastric tumors [7]. When one compares the Laurén and the WHO classification tubular and papillary carcinomas fall within the intestinal type of stomach cancer, whereas signet-ring cell carcinoma and other poorly cohesive carcinomas correspond to the Laurén diffuse type [10][64].

1.3. Goseki classification

The Goseki classification divides GC, based on intracellular mucin production and the degree of tubular differentiation, into four groups: group I: tubules well differentiated, intracellular mucin poor; group II: tubules well differentiated, intracellular mucin rich; group III: tubules poorly differentiated, intracellular mucin poor; group IV: tubules poorly differentiated, intracellular mucin rich. Most studies, which have focused on prognostic significance, did not confirm a prognostic independent value of this system [10][7].
Although current histopathological systems influence endoscopic or surgical choices, they are still insufficient to guide precision treatments for individual patients. Not only new therapies, but a new classification for GC is urgently needed as well.

1.4. The precancerous cascade

The intestinal type of gastric adenocarcinoma is preceded by a sequence of histological lesions (known as Correa´s cascade) with well-defined characteristics: non-atrophic gastritis, multifocal atrophic gastritis without metaplasia, intestinal metaplasia of the complete type, intestinal metaplasia of the incomplete type, dysplasia [16, 17], (Figure **1-2**), ([18], [13], [26]). This precancerous process was described in 1975 by Correa et al. based on observations in Colombian populations at high risk of GC [15]. Following the identification of H. pylori as a causative agent of gastritis in 1983 [67], it was recognized that this process is initiated and sustained by the infection with this bacterium and may last for decades preceding the malignant transformation.

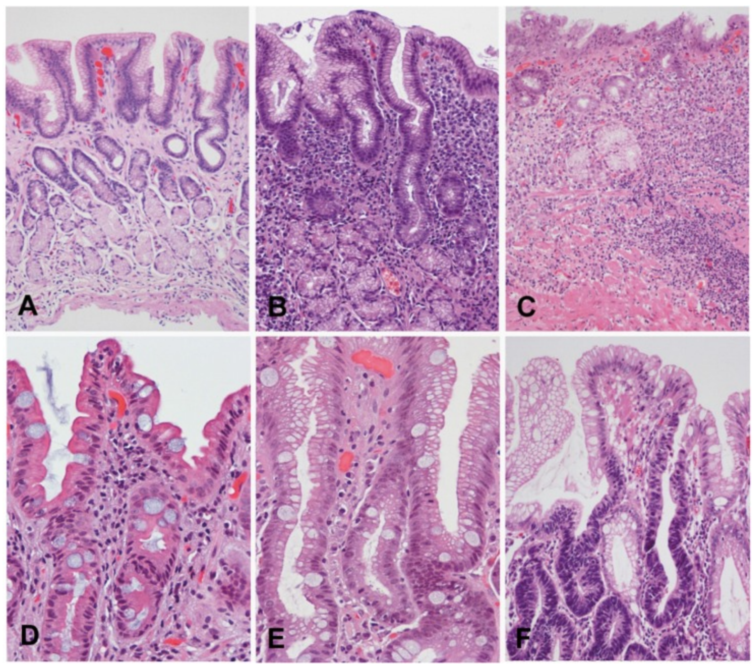

Figure 1-2: Correa's precancerous cascade. A, Normal gastric mucosa. B, Non atrophic chronic gastritis. Abundant inflammatory infiltrate in lamina propria with well-preserved glands observed in the deeper half of the mucosa. C, Multifocal atrophic gastritis without intestinal metaplasia. Marked loss of glands, with prominent inflammatory infiltrate and proliferation of fibrous tissue in the lamina propria. D,Intestinal metaplasia, complete type. Goblet cells alternating with absorptive enterocytes that present well-developed brush border. E, Intestinal metaplasia, incomplete type. Goblet cells alternating with columnar cells that contain mucin droplets of variable sizes. F, Dysplasia. Epithelium with high-grade dysplasia (lower half of the photograph) occurring in a background of incomplete metaplasia (observed in the foveolar superficial epithelium). (HE; original magnification: A-C x100; D-F x200).Images A-Care reproduced with permission reference 60. [14]

Complete intestinal metaplasia is characterized by well-developed goblet cells alternating with mature absorptive cells. Paneth cells may also be observed. Incomplete metaplasia consists of goblet cells alternating with columnar mucous cells containing varying amounts of intracytoplasmic mucin. Dysplasia (glandular intraepithelial neoplasia or non-invasive neoplasia) is a neoplastic lesion limited to the epithelium, without invasion of the lamina propria, and is classified mainly as low grade and high grade. Low-grade dysplasia is characterized by crowded glands lined by columnar cells with hyperchromatic, elongated, and pseudo-stratified nuclei that maintain polarity with respect to the basement membrane. Low-grade dysplasia shows minimal glandular architectural alteration and the cells show mild to moderate atypia. In high-grade dysplasia, dysplastic cells are usually cuboidal rather than columnar, with a high nuclear-cytoplasmic ratio, loss of nuclear polarity, and prominent nucleoli.

China and Colombia are the regions with the highest incidence of gastric adenocarcinoma, [20]. After the stages described above, the stomach tumors that are based on the tubules and originate in the gastric mucosa are called Adenocarcinoma we pass to our focus of study, intestinal or tubular type adenocarcinoma can be classified according to WHO by counting the structural observation of tubules, generally, pathologists perform this count on the area of cancer, but in this investigation the entire image will be made trying to obtain more significant information about the environment, which determines the severity of gastric cancer, it is also possible to observe that the dysplastic processes of pseudostratification or stratification of the epithelial cells that cover the gland, are doubled or tripled, filling with lymphocytes. The infilter carcinoma is the stage more advance [32] begin to form polyp as large than 5cm which is classified with the borrmann stadio then become at metastasis by overcoming the lymph nodes that reach other systems of the human body.

1.5. Glandular and Nuclear Architecture

Some glands exposed to this stress, usually change their shape and their normal size, causing hyperchromatic in nucleus, stromal fibrosis triples, the light or lumen ceases to have an ovoid or regular shape, the superficial epithelium is eroded and foveal glands reflect a tortuous and irregular shape tending to ulcerate easily. Normally, when the architecture of the abnormal mucosa is destroyed, it becomes an irregular and tortuous structure, a key moment to start performing classification analysis [27]. Initially, overgrowth is sought by bacteria, by epithelium, by nucleus, by cell, and if there is mucin generation, since this causes the appearance of piano keys to be lost due to overlapping with another cell; It is also important to analyze the spatial relationship of the gland vs the stroma, since under normal conditions the spatial relationship is 50 % - 50 %, and under abnormal conditions the stroma continues to grow up to 70 % - 90 % of the total cell.
For the pathologist it is of vital importance to obtain a complete and clear image to give a good orientation to the slide, since one can run with the risk of seeing only the superficial part and not being able to see the deep of the epithelium where the formation of a tumor [24]. In this exercise the specialists focus their attention on making a detailed diagnosis of the tumor, thus observing more healthy regions within the same biopsy.

When the nucleus already reaches 80 % of the cell, it is where the tumor is poorly differentiated that it is growing rapidly, the interstitium is minimal, and should be something that can be seen easily, but the cells are back to back preventing the comparison, all tumor gland cells are packed like mitosis cells, there is a lot of chromatin and tumor apoptosis cells dividing very fast.

We have a series of substances or situations that make these mechanisms that control this cell cycle, develop a genetic variation, abandoning the normal cycle of execution, affected by epigenetics influenced by the environment.
The body glands produce hydrochloric acid by decomposing the food, whereas the glands of the antrum produce protective substances such as mucus, which protect the normal structure of the stomach from the acid produced by the corporal glands [23], it is clear that the mucosecretory glands that have inflammatory cells are normal in that region since they are also generally located in blood capillaries fibroblasts.

1.6. Criteria to Give a Tubular Differentiation

For the pathologist it is important to have the complete image of the biopsy to know in the context and the distribution of the glands and to know if it is talking about a corporal gastric mucosa or an antral gastric mucosa. See figure **1-3** and **1-4** the difference between an antral biopsy and a body biopsy.

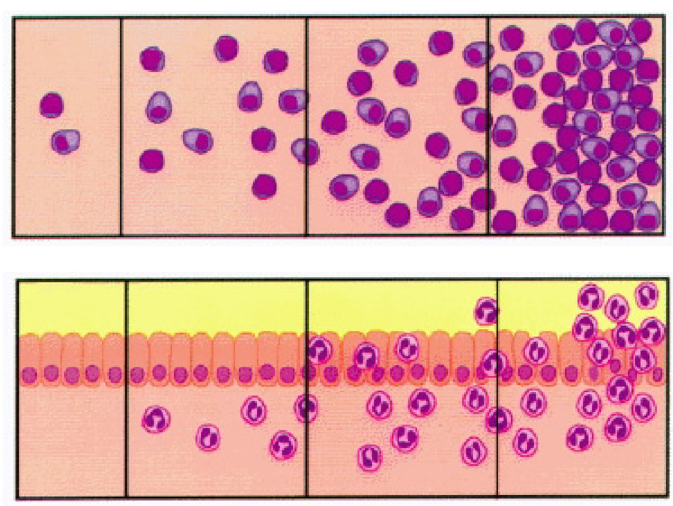

Figure 1-3: Visual scales of inflammatory infiltrates, grades 0 - 1 - 2 - 3 of the Sydney System renewed, 1994. [42]

Several studies show that the work of a pathologist is to identify abnormal regions that influence cell death, also called tumors, and it is on these regions where the tumors can be classified according to their cell density, World Health Organization (WHO)[36] proposes to classify glandular density into three differentiation grades:

- Grade 1 (Well differentiated): greater than 95 % of well-formed cells.

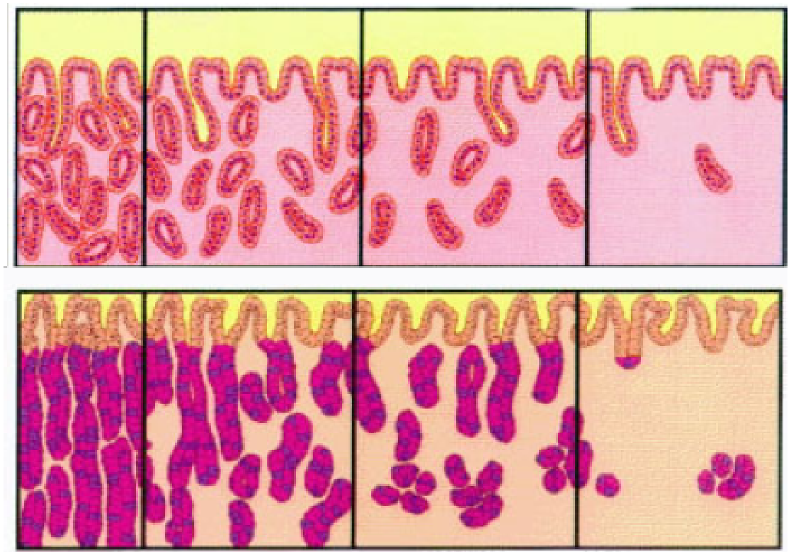

Figure 1-4: Body mucus, Antral mucosa. Visual scales of the degree of atro to antral mucosa and body 0 - 1 - 2 - 3 (normal / atro to mild, moderate, severe) of the renovated Sydney System, 1994 (reproduced with the permission of Prof. RM Genta).

- Grade 2 (Moderately differentiated): between 50 % and 95 % of well-formed cells.
- Grade 3 (Poorly differentiated): less than 49 % of well-formed cells.

To perform this differentiation, the tumor is usually already in a state of adenocarcinoma and it is just at this moment that they begin to analyze and to generate different subjective readings among pathologists since this appreciation is quite appreciative and truly subjective to the experience of each pathologist. Therefore, it was necessary to take into account basic concepts of both gastric mucosal structures and cellular structures, to understand the changes in the stroma and cell parenchyma, and therefore observe losses in the structure between each cancer state.

An image tells the whole process that has a state so far, that is, if it is diagnosed as dysplasia the image of the biopsy evidences all stages antecedent to it in the same or in different regions; [56] The first thing is to determine the area where the existence of cancer is perceived, then determine what type of cancer it is, and if it is tubular to determine if it is in any of the differentiation mentioned by the WHO percentage of well-formed tubules.

According to the conference in Vienna and Padova [19] to achieve operation analyze how gastric biopsies should be reported, it was decided thanks to the consensus to define as grade 1 the normal state, grade 2 to the presence in amatoria, grade 3 to the difficulty to differentiate between inflammatory and tumoral, grade 4 to tumor and grade 5 to fully developed tumor, grade 5 is only used when all stomach is removed, for biopsies grade 4 is understood as the maximum degree of the tumor. Adachi et al. [1] they showed that histology type is important for estimating the tumor progression and outcomes of patients with gastric carcinoma. In addition to the depth of wall invasion and status of lymph node metastasis, histologic type, including well or poorly differentiated type, should be evaluated in the management of GC.

1.7. Research Problem

Quantification of glands might help to determine the aggressiveness of cancer; however, as it mainly relies on the visual interpretation of a pathologist, it unavoidably leads to a certain level of subjectivity[34], consequently, automated tools could help to standardize diagnostic processes. The underlying hypothesis of this investigation is that the use of visual characteristics that model the geometry of the glands and their cellular properties could help to determine the disease' s aggressiveness.

1.8. Contribution

A main contribution of this manuscript are

- local and contextual features that describe glandular nuclei
- A gland description in terms of the previously nuclei features
- A complete approach to predict survival after a year in GC patients. This estimation was correlated with the degree of the disease in spite of the small data set for evaluation.

2 An automatic segmentation of gland nuclei in gastric cancer based on local and contextual information

Presented on ”Biomedical Information Processing and Analysis - A Latin American perspective”. SAMBA: SIPAIM, MICCAI BIOMEDICAL WORKSHOP, March, 2019.

Analysis of tubular glands plays an important role for gastric cancer diagnosis, grading, and prognosis; However, gland quantification is a highly subjective task, prone to error. Objective identification of glans might help clinicians for analysis and treatment planning. The visual characteristics of such glands suggest that information from nuclei and their context would be useful to characterize them. In this paper we present a new approach for segmentation of glandular nuclei based on nuclear local and contextual (neighborhood) information. A Gradient-Boosted-Regression-Tree classifier is trained to distinguish between glandular nuclei and non-glandular nuclei. Validation was carried out using a dataset containing 45.702 annotated nuclei from 90 1024x1024 fields of view extracted from gastric cancer whole slide images. A Deep Learning model was trained as a baseline. Results showed an accuracy and f-score 5.4 % and 23.6 % higher, respectively, with the presented framework than with the Deep Learning approach.

2.1. Introduction

Gastric cancer (GC) is among the most diagnosed cancers and the second most frequent cause of cancer-related death worldwide [40]. Geographically, the highest incidence of GC is in Asia, Latin America, and the Caribbean [[**?**],[**?**]]. In Colombia, GC is the first cause of cancer-related death, representing a 15 % of all cancer deaths, with a high incidence in the Andean zone, especially in the departments of Nariño, Boyacá, and Cundinamarca. Currently, it is considered a major public health problem that has generated an expense of more than 47 million USD in five years [11].

GC comprises several kinds of lesions with different severity grades. From such lesions, adenocarcinoma is the most common, representing more than 90 % of all GC [44]. Characterization and quantification of the adenocarcinoma might establish plausible chains of events that improve the disease understanding and reduce its mortality rates. Diagnosis is usually reached by an endoscopic biopsy of the stomach which is processed and analyzed by pathologists who determine the degree of malignancy [69]. One of the most common approaches to identify and grade gastric adenocarcinomas is by identifying and estimating the density of glands. Low-grade lesions are characterized by the presence of well/moderately differentiated glands (Figure **2-1**-a). In high-grade lesions, glands are highly irregular and poorly differentiated (Figure **2-1**-b) [[49], [44]]. Identification of glands plays an important role not only in diagnosis but also in establishing some prognosis [49]. An accurate quantification is therefore essential for both the decision making flow and the treatment planning. Unfortunately, this process has remained highly subjective and prone to error. In this context, automatic measures may contribute to identify tubular glands on GC samples.

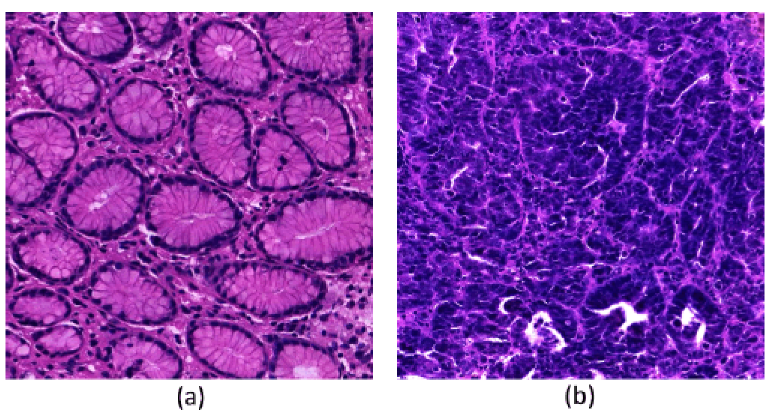

Figure 2-1: Representative images of Hematoxylin-Eosin stained tissue from gastriclesions. a)Well-differentiated glands, b)Poorly-differentiated glands.

This work introduces an automatic strategy that exploits nuclear local and contextual information to identify gland nuclei in fields of view (FoVs) extracted from gastric cancer whole slide images (WSIs). The present approach starts by automatically segmenting nuclei with a watershed-based algorithm [65]. Each nucleus is then characterized by two types of features: first, its own morphological properties (size, shape, color, texture, etc.), second, its neighbor nuclei features within a determined radius. Such features are used to train a Gradient-Boosted-Regression-Tree (GBRT) classifier to differentiate between gland-nuclei and non-gland-nuclei. Unlike other state-of-the-art methods, any feature in this approach exploits nuclei relative information, i.e., any nucleus information is always relative to how such feature is with respect to its surrounding nuclei. This strategy is compared with a Deep Learning (DL) model that was trained to identify gland-nuclei. This DL model receives as input patches from WSI and outputs probability maps that are thresholded. A watershed-based algorithm segments then the binary output map and splits the connected/overlaid cases to set the final candidates.

2.2. Methodology

2.2.1. Preprocessing: Nuclei Segmentation

A watershed-based algorithm [65] is applied to segment nuclei, generating a mask with the position of each nucleus. Each detected nucleus is then assigned to the class either glandular nuclei or non-glandular nuclei (See Figure **2-2**).

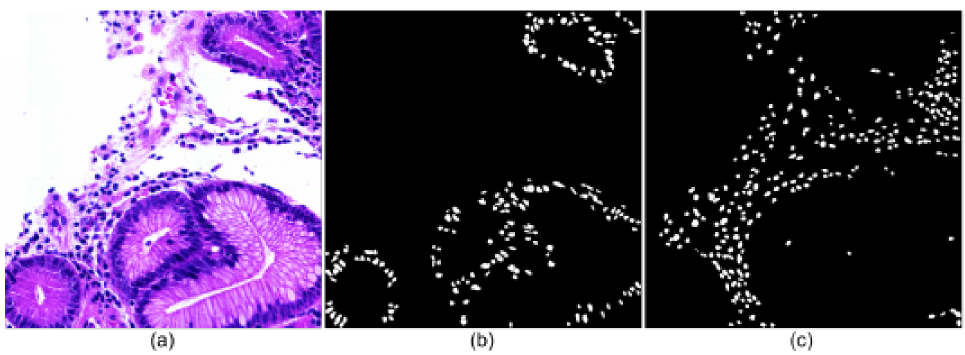

Figure 2-2: Description of the nuclei segmentation. a)Original image, b)Glandular nuclei mask, c)Non-glandular nuclei mask.

2.2.2. Nuclear Local and Contextual Information (NLCI)

In H&E images, glandular nuclei are generally distinguished from other cell nuclei by their orientation, color, oval shape, eosinophilic cytoplasm, and proximity to other similar nuclei. For this reason, after nuclei were segmented, a set of low-level features were extracted, including shape (nuclei structural area, ratio between axes, etc.), texture (Haralick, entropy, etc.), and color (mean intensity, mean red, etc.). Each nucleus was represented by this set of local features. Additionally, for each nucleus, a set of circles with incremental radio of k = dL x 10; dL x 20; dL x 30 pixels were placed at the nucleus center (begin dL = 20 pixels the averaged nuclei diameter), aiming to mimic a multi-scale representation. Finally, a set of regional features was computed within each circle and used to characterize each of the segmented nuclei. These features measure the neighborhood density and relative variations in color, shape, and texture.

A set of 57 local and contextual features were extracted from each image nuclei and the 33 most discriminating characteristics were selected by distribution analysis and Infinite Latent Feature Selection (ILFS) algorithm [57]. A GBRT classifier [30] was then trained to differentiate between the glandular nuclei and non-glandular nuclei classes. Specifically, we used an adapted GBRT framework[6] which emphasizes the minimization of the loss function.

2.2.3. Baseline

State-of-the-art

The baseline corresponded to a state-of-the-art deep learning approach known as Mask Region-based Convolutional Neural Network (R-CNN). This modification of the Fast R-CNN algorithm [31] has been used in the Kaggle 2018 Data Science Bowl challenge for identifying wider range of nuclei across varied conditions [35]. It uses a deep convolutional network with a single-stage training and a multiscale object segmentation. Mask R-CNN outputs an object detection score and its corresponding mask [31].

DL model

The DL model was trained using a set of patches extracted from the FoVs. The positive class patches correspond to the area covered by the bounding box of each gland nucleus while the negative class patches were taken from the background, i.e., regions with non-gland nuclei. Aiming to increase the number of training samples, different transformations (e.g., rotation and mirroring) were applied to the patches. Model training was carried out using a total of 20 epoch cycles with 100 steps each.

Figure **2-3** presents the architecture of a trained DL model for the exploratory stage. A random extraction of a Region of Interest (RoI) is performed. This RoI is projected to a convolutional network that generates a feature map. These features are introduced to the RoI pooling layer for further processing. At the last stage, fully connected layers generate the desired outputs, including the gland nuclei candidate bounding box and mask.

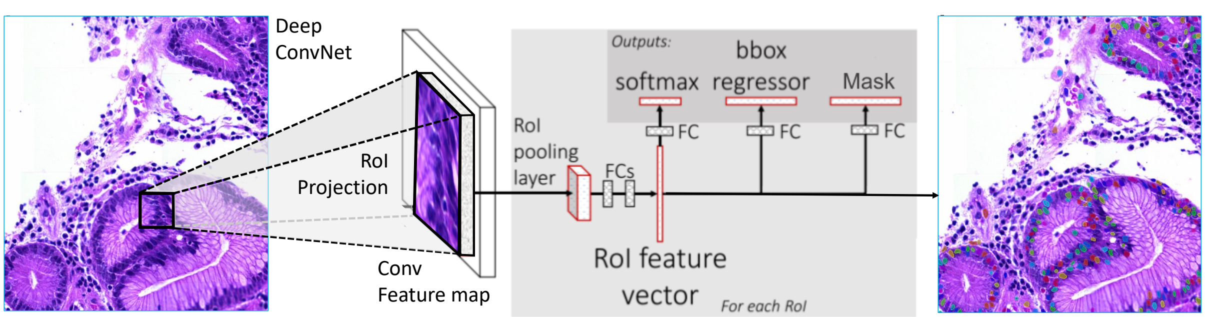

Figure 2-3: Mask R-CNN work flow. Figure extracted and adapted from [31]

2.3. Experimentation and Results

2.3.1. Dataset

The dataset consisted of 90 FoV of 1024 x 1024 pixels at 40x extracted from a set of H&E WSI taken from 5 patients who were diagnosed with GC. The WSI were provided by the Pathology Department of Universidad Nacional de Colombia. A total of 45.702 glandular nuclei were manually annotated, being 12.150 glandular nuclei while the remaining 33.552 corresponded to other structures (non-glandular nuclei).

2.3.2. Experiments

Two experiments were carried out. The first attempted to classify between glandular nuclei and non-glandular nuclei using the NLCI approach. A Random Cross validation method with 10 iterations was used. At each iteration, 70 % of the whole set of FoV was used to train the GBRT classifier and the remaining 30 % was used to test the trained model. Finally, the measured performances along the 10 iterations were averaged.

The second experiment aimed to identify glandular nuclei using the DL model. For this purpose, 60 FoV were used to train the model and the remaining 30 for testing. In this case, glandular nuclei detection was assessed based on the number of detected nuclei centroids that correctly overlapped with the ground truth nuclei, judged as correct when centroids were within one nuclear radius.

2.3.3. Results

Table **2-1** presents different performance metrics for both assessed approaches. NLCI achieved an accuracy of 0.977 and an F-measure of 0.955, while R-CNN yielded corresponding accuracy and F-measures of 0.923 and 0.719, respectively. For the qualitative results, Figure **2-4** shows the resulting gland nuclei segmentation from both approaches, where R-CNN generates its own masks of single gland nucleus presented by individual colors.

Metrics	NLCI	R-CNN
Accuracy	0.977	0.923
Precision	0.959	0.585
F-score	0.955	0.719

Table 2-1: Comparative measurements for both approaches.

2.4. Conclusions

In this chapter, two different approaches to automatically detect glandular nuclei on gastric cancer images were presented and compared: a model based on nuclear local and contextual information and a DL

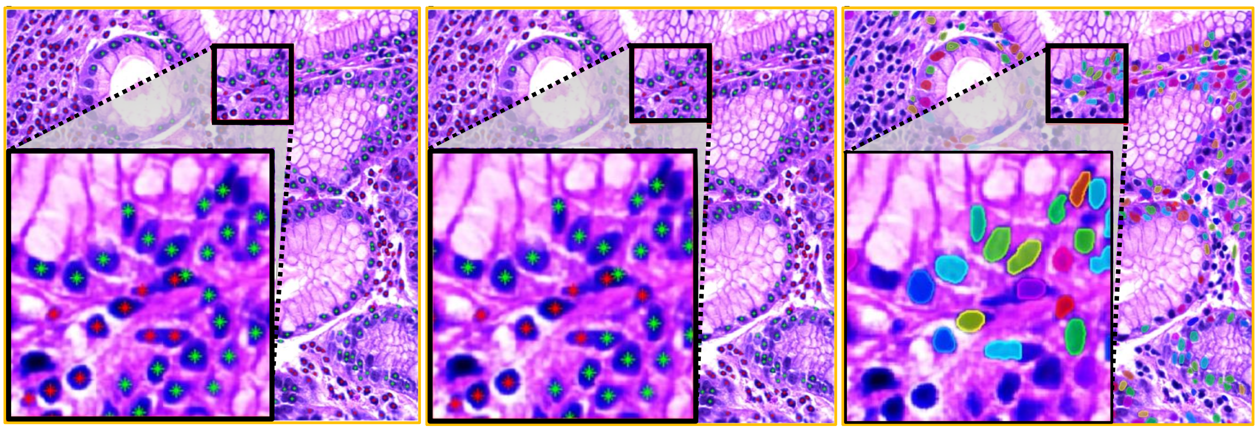

Figure 2-4: Gland nuclei Segmentation showing, a)The ground-truth label , b)NLCI, and c)R-CNN with each gland nuclei candidate individually colored.

model. Results demonstrate that local and contextual features are appropriate to describe the structural features of tubular glandular nuclei. Despite the DL model presented good results, this approach requires a powerful/expensive infrastructure, long training times, and huge quantities of annotated data. Due to the lower precision of the model, it indicates the that only local information its taken into account. Future work includes quantification of glands to establish correlation with cancer grade and patient prognosis. [38]

3 A method to detect glands in histological gastric cancer images

Presented on the 14th "International Symposium on Medical Information Processing and Analysis" (SIPAIM) conference, Mazatlán, Mexico, 2018. Published in: SPIE Digital Library, 28 December 2018, Mazatlán, Mexico.

Automatic detection and quantification of glands in gastric cancer may contribute to objectively measure the lesion severity, to develop strategies for early diagnosis, and most importantly to improve the patient categorization. This article presents an entire framework for automatic detection of glands in gastric cancer images. This approach starts by selecting gland candidates from a binarized version of the hematoxylin channel. Next, the gland's shape and nuclei are characterized using local features which feed a Random lo Cross validation method classifier trained previously with manually labeled images. Validation was carried out using a data-set with 1.330 annotated structures (2.372 glands) from seven fields of view extracted from gastric cancer whole slide images. Results showed an accuracy of 93 % using a simple linear classifier. The presented strategy is quite simple, flexible and easily adapted to an actual pathology laboratory.

3.1. Introduction

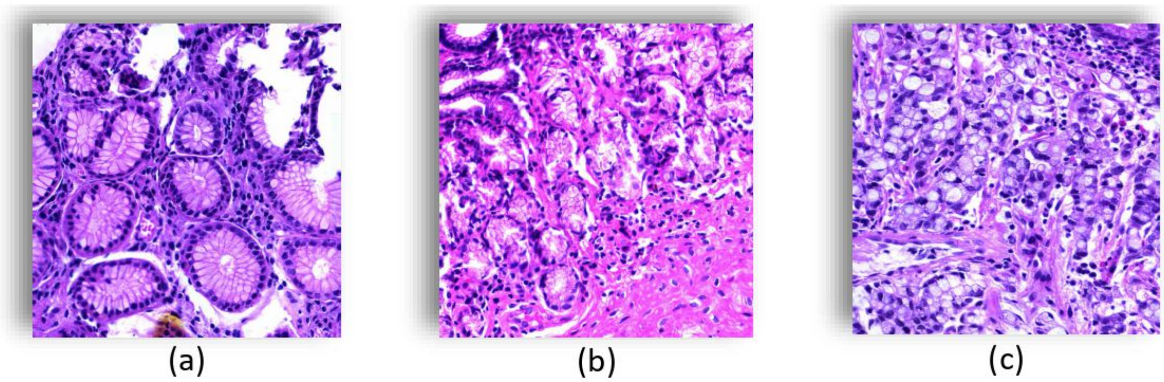

Figure 3-1: Description of the lumen segmentation. a) Well differentiated (More than 95 % of the tumor is composed of Healthy glands), b) Moderately differentiated (50 % to 95 % of the tumor is composed of Healthy glands) , c) Poorly differentiated (49 % or less of the tumor is composed of Healthy glands).

Gastric cancer (GC) remains an important public health problem. [39] About one million new cases have been yearly reported, becoming the fourth most common cancer and the second leading cause of cancer deaths worldwide. [63] This cancer has been related with the infection of Helicobacter Pylori [66] yet its physiopathogenesis is still unknown and many other factors have been correlated with. Several studies have shown that characterization and quantification of GC may improve understanding about the chain of events that triggers the disease.[69]

Adenocarcinoma is the most common type of GC. The most common strategy to classify adenocarcinoma is to estimate the degree of gland differentiation. The world health organization (OMS) established [36] that grades 1 and 2 conserve most morphological gland properties, basically shapes and sizes. Higher degrees are characterized by more irregular and hyperchromatic structures (Fig.**3-1**). Gland identification supports diagnosis, categorization and prognosis of the patient. Therefore, precise quantification is essential for both decision making and treatment planning. Unfortunately, so far, this process has remained highly subjective and prone to errors. [21]

3.1.1. State-of-the-art

Despite the importance of detection and quantification of glands in gastric cancer, few methods are reported in the literature. Ficsor et al.[29] analyzed 79 cases of gastrointestinal biopsies stained with hematoxylin and eosin (H&E). They determined a set of heuristic cytometric parameters and use them to classify cases as normal mucosa, gastritis, or adenocarcinoma. Three nonparametric methods established a general correlation of 86 % between the number of glands and the particular pathology. Similar works [51] have explored detection of glands in breast tissue by integrating image information at three scales: (1) low level or pixel values, (2) high level or object detection, and (3) relationship between structures. A Bayesian classifier assigns then a probability class to each pixel. Doyle et al. [22] performed segmentation of glands in prostate cancer. The authors used morphology priors to achieve gland segmentation. Then, SVM classified tissue patches containing benign epithelium and the prostate cancer degrees 3 and 4. Results suggested quantitative analysis of the gland morphology may play a significant clinical role in distinguishing different prostate cancer degrees.

3.1.2. Contribution

This chapter presents a novel automatic strategy that exploits morphological information from lumens and nuclei to detect glands in gastric samples. The method is simple, computationally non-expensive, and independent of parameter tuning. The main contribution of this strategy is a set of discriminating visual characteristics that model the geometry of actual glands and their cellular properties.

3.2. Methodology

Gland candidates are firstly repaired in a binary version of the hematoxylin channel (panel (c) of figure **3-2**). The original image has been previously separated into the two hematoxylin and eosin channels.[65] Big objects in the binary image are then selected, dilated and superimposed with the original image (panel (d) of figure **3-2**). On the other hand, nuclei are segmented by a watershed method (panel (e) of figure **3-2**) and the cells intersecting the gland candidate (panel (f) of figure **3-2**) shall be used to characterize the gland candidate. Overall two types of features will be extracted, geometric from the gland candidate in panel (c) and cellular from the cells surrounding the gland, as illustrated in panel (f) of figure **3-2**. This approach, compared to other cutting-edge methods, is simple, very easy to use and implement while it presents good performance, fast training times and precise results. The whole method is further presented hereafter.

3.2.1. Preprocessing

The workflow of the proposed method is composed of the following stages. First, the digitization of a significant set of cases of previously diagnosed GC, existing in the sheet banks of the district network of hospitals. Then, assembling each one of sheets digitization, this process consists in the reconstruction of virtual plate that is stored in a server in (http://cimalab.unal.edu.co/microscopio/viewer?wsi=002-14reader=jpeg), in this case, the corresponding plates of H&E and a process that will allow the overlapping of the sheets.

3.2.2. Candidate glands

Provided that glands are characterized by a lumen surrounded by a layer of cells, the whole method is focused on detecting this type of structure. A first step is a color deconvolution [46] to obtain the H&E (hematoxylin and eosin) channels. Then, a Gaussian filter is used to smooth the hematoxylin channel and a threshold is set at 90 % of the max intensity to detect white big regions, see fgure **3-2**(b). The number of candidate glands is reduced by applying an initial erosion with a disk-structuring element of 9 pixels (about half diameter of the smallest nucleus), followed by filling eroded lumens and removing small objects (less than 10 pixels), see figure **3-2**(c). Finally, a set of regions (candidate glands) are obtained.

3.2.3. Nuclei surrounding the candidate gland

Nuclei were segmented by using the method proposed by Veta et al [65]. This starts with a color deconvolution of the H&E image to isolate the hematoxylin channel.[37] Then, a fast radial symmetry transform is applied to find nuclei related markers and from there the watershed algorithm detects nuclei as connected components (figure **3-2** panel (e)). The candidate gland obtained from the binary image is then dilated by a disk-structuring element of 60 pixels (maximum diameter of the largest nucleus) and the set of nuclei overlapping this regions is then selected (figure **3-2** panel(f)) for gland characterization.

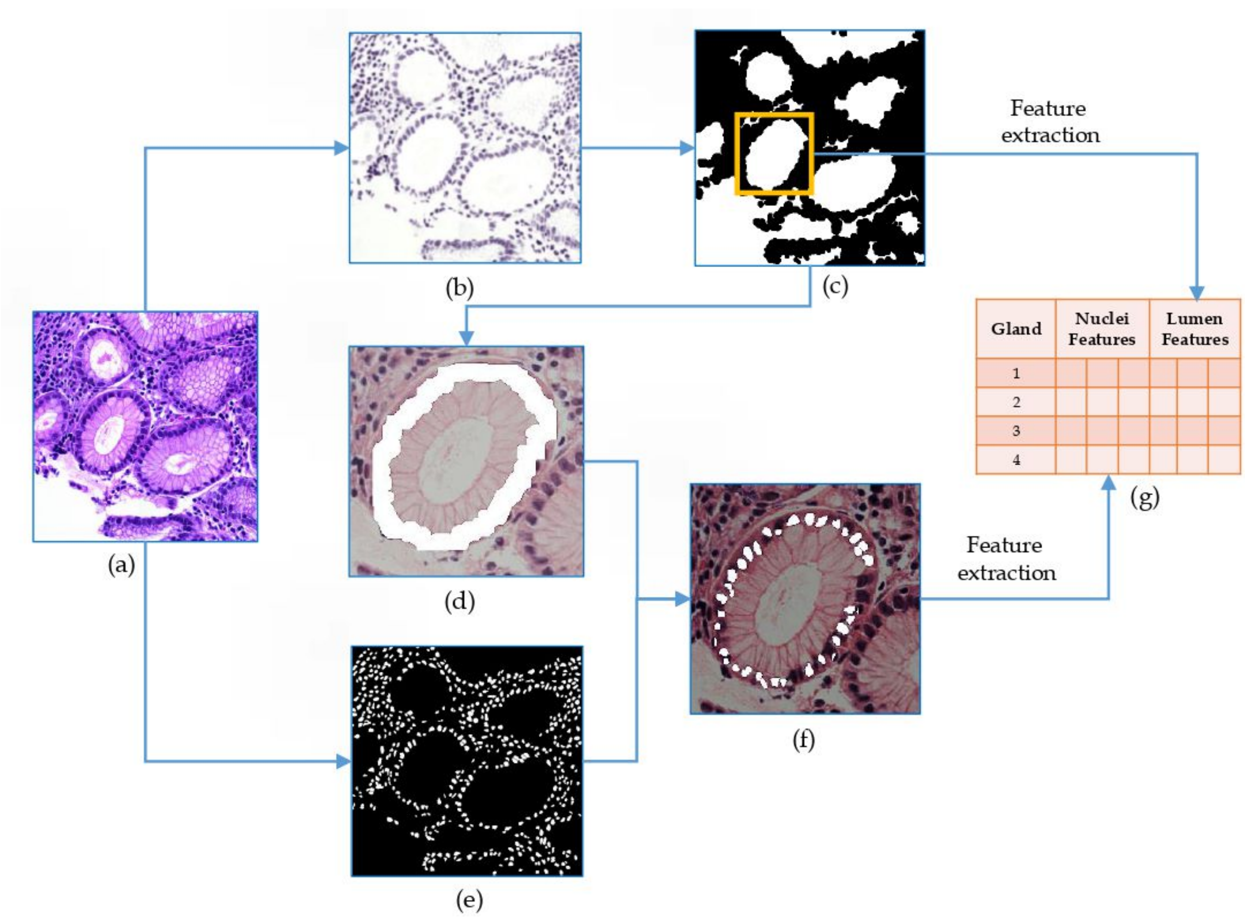

Figure 3-2: The process to obtain candidate glands: a) Original Image, b) Hematoxylin Image, c) Candidate glands d)Dilation, e) Nuclei Detection, f)Nuclei surrounding the candidate gland, g) Feature matrix.

3.2.4. Feature extraction

Each candidate region is then described by a set of nuclei characteristics extracted from the nearest nuclei neighbors of the dilated version of the gland candidate boundary. Different features are extracted from this set of cells, namely shape (area, perimeter, texture, orientation), color (intensity and entropy of the channels) and distance to other nuclei.

A similar set of features is extracted from the lumen candidate, including the shape (lumen area, relation between lumen axes), homogeneity, main orientation, Zernicke moments,[59] Haralick Textural Features[22] and color (average intensity, mean and variance of the red channel).

3.3. Experimentation and Results

3.3.1. Dataset

1.330 Fields of View (FoV) of 1024 x 1024 pixels at 40x was extracted from a set of H&E Whole Slide Images (WSI) from 7 patients diagnosed with adenocarcinoma (n = 3), gastritis (n = 2), and metaplasia (n = 1) by two pathologists. The WSI were provided by the Pathology Department of Universidad Nacional de Colombia. FoV with glands were manually annotated by at least one pathologist and used to train the different models. A total of 11.689 structures were annotated, being 2.372 glands and the remaining 9.317 corresponded to other structures.

3.3.2. Results

The introduced method was validated by classifying the detected candidate regions as either glands or non-glands. A Random Cross validation method with 10 iterations was used. At each iteration, 70 % of the whole set of FoV was used to train a Gradient boosting tree (GBT) classifier [6] and the remaining 30 % was used to test the trained model. The GBT was trained setting a shrinkage factor of 0.1, a subsampling factor of 0.5, and a max tree depth of 2. Finally, the measured performances along the 10 iterations were averaged. Figure **3-3** presents the Receiver Operating Characteristics (ROC) curve corresponding to the classification task and Table **3-1** presents some performance metrics. Results show that the presented approach yielded an accuracy of $\tilde{9}0$ % and an F-score of 72 %, suggesting that the introduced approach might be suitable for the identification of glands.

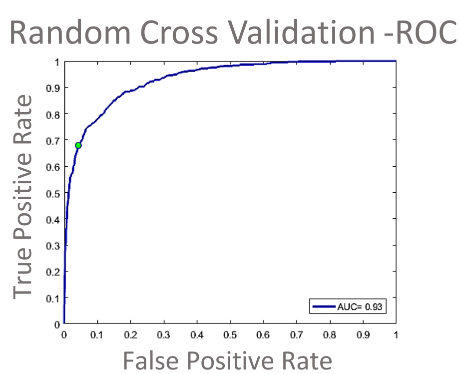

Figure 3-3: ROC

Metrics	Menu Values	Sid Values
Area Under The ROC	0.933	0.003
Accuracy	0.9	0.005
Sensitivity	0.638	0.02
Specifity	0.968	0.005
Precision	0.836	0.021
F-score	0.723	0.015
Geometric mean	0.768	0.012

Table 3-1: Measurements

3.4. Conclusions

In this article, a method for automatic identification of glands in gastric tissue samples was presented. The introduced approach exploits morphological information from both lumens and nuclei of gastric glands. Results suggest that the introduced approach is suitable for identification of glands in gastric tissue. This approach could contribute to objectively quantify glands and thereby grade GC lesions.

4 Searching Histology Patterns in Gastric Glands for predicting Gastric Cancer Survival

Paper accepted by SPIE Digital Library. (SPIE Paper Number 11320-42 Acceptance and Manuscript Information)

This article presents an entire framework for analyzing survival-related gland features in gastric cancer images. This approach builds upon a previous automatic gland detection, which partitions the tissue into a set of primitive objects (glands) from a binarized version of the hematoxylin channel. Next, gland shape and nuclei are characterized using local and contextual features that include relationships between color or texture from glands and nuclei ($5{,}120$ *features). A mutual information max-relevance-min-redundancy (mRMR) approach selects hundred features that correlate with patient survival "survival vs not survival (first year)". Finally, ten statistically significant features (test* t*-student,* $p < 0{,}05$*) were used to set a "one-year"survival. Evaluation was carried out in a set of fourteen cases diagnosed with pre-cancerous gastric lesions or cancer, under a leave-one-out scheme. Results showed an accuracy of 78.57 % when predicting the patient survival (less or more than a year), using a QDA Linear & Quadratic Discriminant Analysis. This approach suggests there exist morphometric gland differences among cases with gastric related pathology.*

4.1. Introduction

Gastric cancer (GC) incidence and mortality have been reduced over the past 70 years [61]. Despite a recent decline, worldwide it is still the fourth most common cancer and the seventh leading cause of cancer-related death [53],[54]. Geographically, the highest GC incidence has been reported in Japon, Latin America, and the Caribbean [43, 28]. In Colombia, GC is the first cause of cancer-related death, representing a 15 % of all cancer deaths, with a high incidence in the Andean zone, especially in the departments of Nariño, Boyacá, and Cundinamarca. Currently, it is considered a major public health problem whose economic burden has reached the 47 million USD in the last five years [12].

A GC diagnosis and stratification [34] is achieved by examining a biopsy tissue under a microscope[9]. This mainly relies upon certain level of expertise [33], a limited resource in actual pathology laboratories. Overall, such diagnosis is not exempt of an inevitable observer bias and subjectivity. In this context, an automatic characterization of gastric glands may objectively support diagnosis and lead to devise more accurate indexes to predict the disease evolution. [60, 25].

4.1.1. State-of-the-art

To the best of our knowledge, few investigations have aimed to determine survival in GC populations. Williams et al.[68] integrated multiple databases of patients diagnosed with GC including pathological, clinical, surgical and survival information. They applied a Machine Learning methodology to characterize subgroups of patients with gastric cancer by exploring all relationships between patient descriptors and systematically extracted over 450,000 logical associations. A subset of more than 1000 associations identified possible disease risk markers. Oh et al.[52] developed an automatic model to predict survival outcomes for patients with GC using a recurrent neural network (RNN). This study enrolled 1.243 cancer patients. Results showed a ROC AUC of 0.81 for the survival recurrent network (SRN) data test.

4.1.2. Contribution

A main contribution of this work is an automatic characterization of gastric glands together with a set of features that might be associated with the disease aggressiveness. This set of characteristics correlates with the survival time ($\pm 1 year$) in a group of patients with gastric pathology. Furthermore, these discriminatory features are used in a classification task to build a model for predicting the patient survival time.

4.2. Methodology

A set of morphological and textural features are extracted from gastric glands automatically detected and their nuclei [2]. Using a max-relevance-min-redundancy (mRMR) criterion these features are reduced from about six thousand features to barely a hundred. These features are then statistically assessed to identify the ones that better express differences and these ones are then used to train Quadratic Discriminant Analysis (QDA) classifier.

4.2.1. Characterization of gastric glands

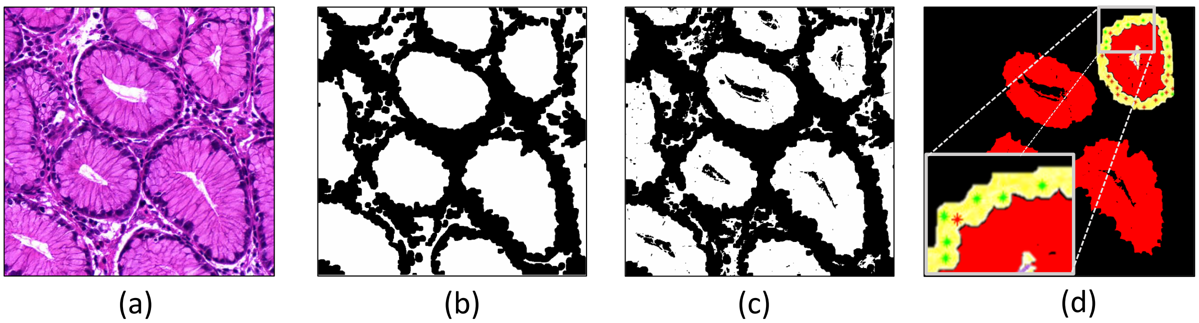

Figure 4-1: Proposed Methodology: a) Original Image, b) Gland binary mask, c) Gland candidates, d) Gland Nuclei

A coarse Gland binary mask is firstly constructed [2], as illustrated in panel (b) of figure **4-1**, by thresholding the hematoxylin channel, previously determined by a color deconvolution technique [46]. The original image in panel (a) is then thresholded and subtracted from a version of the image in panel (b) whose impulse noise has been filtered out by specific morphological operations, i.e., erosion and the area operator, which switches the binary value of all zones whose areas are smaller than a given value, see panel (c). Every gland intersecting the image border is excluded from this analysis.

A next step is the search of gland nuclei, a process starting by segmenting all nuclei and determining which of them belong to the gland. For doing so, gland candidates previously found are dilated by the maximum diameter of the largest nucleus (disk-structuring element of 60 pixels) and nuclei are evaluated by a Gradient-Boosted-Regression-Trees model to distinguish between gland-nuclei and non-gland-nuclei [5], see panel (d). This classifier was trained using a set of $45{,}702$ manually annotated gland nuclei, characterized by shape, texture and color features presented in table **4-1**. This characterization also includes local and neighborhood analyses: while a local feature decomposes nuclei in terms of their geometric or physic characteristics, the neighborhood properties aim to capture nuclei in terms of their environment and population features. The neighborhood analysis is basically a spatial exploration of the region surrounding a nucleus and for doing so a set of circles with incremental radii of $k = dL \times 10$ pixels is placed at any nucleus center, starting with the average nuclei diameter ($dL = 20$ pixels) until $dL = 50$ pixels. Neighborhood features are computed from the nuclei inside the circles, in this case the nuclei and cytoplasm characteristics shown in table **4-1**. A nucleus is represented by a vector with 52 features which correspond to such local and neighborhood descriptions.

A gland description is achieved by averaging these 52 features among the whole set of gland nuclei. The nuclei gland is then described by a vector with 104 characteristics composed of 52 feature averages, 52 standard deviations of these features and 24 gland characteristics.

Features	Gland Nuclei			Gland
	Local	Neighborhood		
		Nuclei	Cytoplasm	
Shape	Area, perimeter, eccentricity, longness equiv-diameter, ratio between axes, Angle between axes, orientation, oval shape	Zernike moments, proximity to othe similar nuclei	Sum Inverse Distance of Nuclei, quantitative variance, areas and eccentricity variance, diameter variance, orientation variance, longness, ratio area	Area, relation between axes, perimeter, equiv-diameter, eccentricity, Zernike moments, mayor axes orientation.
Color	Max. intensity, Min. intensity, Mean intensity, Mean of red channel	Median red channel	Ratio Min. Intensity, Ratio RGB, Ratio Mean. Intensity, Ratio Red, Ratio Max. Intensity	Mean of red channel, Mean intensity, Mean intensity and variance of red channel, Max. intensity and Min. intensity, red entropy, EdgeMedIntensity
Texture	Entropy intensity, entropy of red channel	Entropy, Haralick features	Eosinophilic cytoplasm, Haralick features	Haralick features, entropy red channel, entropy intensity

Table 4-1: Features extracted from gland, gland nuclei and their cytoplasm.

So far exploration has been devoted to nuclei (local features) and neighborhood characteristics which are spatially extracted from a series of circles placed at the gland nuclei. Notice these features highly depend on the gland size and shape, two characteristics probably result of the biological sample treatment. These features in consequence are not absolute measures and therefore the relevant relative relationships were found out by a selection process. For doing so, an exhaustive computation of every relation between features was carried out, v.g. $\frac{\text{mean nuclei texture}}{\text{whole gland texture}}$, making the original 104 nuclei and 24 gland features are mapped to a new vector of $4,992$ relations, always respecting any relation is set between nuclei and gland characteristics. Finally, the gland feature vector corresponds to the concatenation of the original vector and the previously described relation feature vector, for a total of $5,120$ dimensions which is then pruned by the selection process.

A feature selection is performed in two steps: First, Minimum Redundancy Maximum Relevance (mRMR) approach is used to select relevant features [55] by minimizing the mutual information between features and maximizing the join probability of the selected features between classes "survival vs not survival (first year)". Afterwards, selected features correspond to those showing significant statistical differences (test t-student, $p < 0{,}05$) between the two survival groups.

Finally, the selected gland features are used to train a quadratic discriminant analysis (QDA) and obtain a model to predict patient-survival time.

4.3. Experimentation and Results

4.3.1. Dataset Acquisition

This data-set was composed of 14 cases, description shown in table **4-2**. Due to the high variability between individuals, applied inclusion and exclusion criteria are reported in table **4-3**.

Data-set	Men	Women	Average Age	Min age	Max age
14	9	5	58	22	87

Table 4-2: Data-set Acquisition

Inclusion criteria
Patients older than 18 years of age.
Histopathological diagnosis of acute and chronic gastritis.
Intestinal metaplasia
Gastric cancer in-situ or advanced.
Without prior treatment and newly diagnosed.

Exclusion criteria
Women in gestation or lactation period
Underweight and/or malnutrition
In chemo-radiotherapy
Surgically operated in the last year
With autoimmune disease
Who have received blood transfusions in the last 6 months
Who have received treatment with antibiotics in the last 2 months
With recent infectious processes less than 2 months

Table 4-3: Inclusion and exclusion criteria

Gastric Cancer WSI were provided by Universidad Nacional de Colombia [50] and training glands were annotated by one expert pathologist. Samples were obtained and digitized with a signed "informed consent" that followed the Helsinki protocol [4]. Table **4-4** shows the survival-time of 14 patients, 8 survived less than one year and the remaining 6 cases survived more than one year. The complete data set corresponds then to these 14 cases in which the survival-time was reported.

4.3.2. Results

A total of 638 Fields of View (FoV) of 1024×1024 pixels at $\times 40$ magnification were extracted from a set of H&E WSI, digitized from the 14 patients diagnosed with adenocarcinoma (n=9), gastritis (n=2), and dysplasia (n=3). From these FoV's 2.076 structures were found out and characterized by the model. The dimensionality reduction was achieved by a Minimum redundancy maximum relevance feature selection, finding the 100 most relevant features. Afterwards, statistical differences are computed to reduce the original set of 100 features to only the 10 most relevant characteristics, which are reported in table **4-6**.

Cases	Diagnostic	Year of death	Years of survival
7	Adenocarcinoma	2014	0
1	Dysplasia	2014	0
2	Adenocarcinoma	2014	1
1	Gastritis	2015	1
1	Dysplasia	2016	2
1	Dysplasia	2018	4
1	Gastritis	2018	4

Table 4-4: All biopsies were taken in 2014, No survival (NS) less than 1 year = 8, Survival (S)more than 1 year = 6.

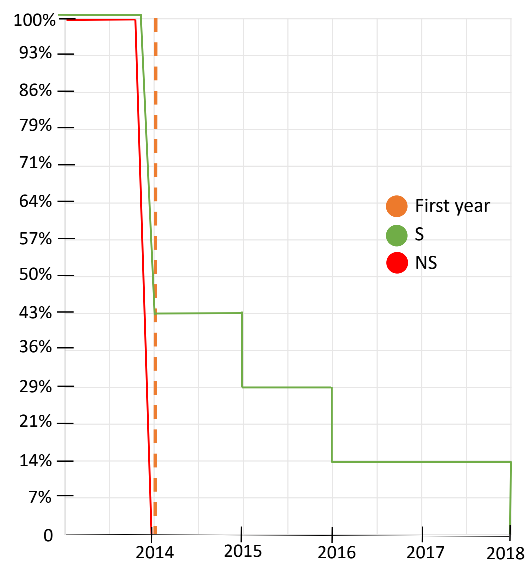

Table 4-5: Survival time - Kaplan meier plot

This last feature selection is used to train a QDA model and predict the survival time of a patient ($\pm 1 year survival time$). The model is trained using a leave-one-case out scheme validation, due to the small number of cases. That is to say, set aside one case for testing and use the remaining 13 GC cases for training, a task repeated 14 times. The final survival label is given by establishing the majority vote of the predictions for all the found glands of the test case. This model demonstrated an accuracy of $78{,}57\,\%$ for the survival prediction task. Additionally, these 10 features are used in a multivariate regression COX model to determine the hazard ratio in each variable of both groups, thereby establishing a risk factor for each of the computed features. This risk ratio, or Hazard ratio, is a relative measure of how relevant a characteristic may be. For instance, for characteristics number 2, 3, 7, 8 and 9 in table **4-6**, a value greater than one suggests that a change in the nuclei texture, cell proliferation per unit of area, the loss of cytoplasm- nuclei relation or nuclei hyperchromatism are linked with aggressiveness of the tumor. All these features have been widely described in most pathology manual as being important to describe aggressiveness.

N	Feature relevant	Relation	Clinical interpretation	Hazard Ratio
1	Nuclei Intensity mean vs glands texture Variance	Color / Texture	Nuclei Hyperchromatism	0.832
2	Nuclei Inverse Difference Moment Std	Nuclei Texture	Nuclei inflammatory changes associated by tumor	1.134
3	SumAverages between nuclei vs SumAverages between glands	Texture / Texture	Lymphoid aggregates	1.047
4	Nuclei Orientation vs glands Area	Orientation / Shape	De-differentiation - Loss of normal structure	0.043
5	Nuclei EntropyIntensity Std	Nuclei Color	Nuclei Heterogeneity	0.038
6	Eosin, Inverse Difference Moment Mean	Cytoplasm Texture	Larger nuclei and glands	0.991
7	Nuclei Info.MeasuresCorrelation 2 mean vs glands Info.Measures Correlation 2	Texture / Texture	Cell proliferation per unit area	1.013
8	Eosin Info.MeasuresCorrelation 2 Mean	Cytoplasm Texture	Loss of cytoplasm- nuclei relation	1.085
9	Eosin, Entropy mean	Cytoplasm Texture	Nuclei Hyperchromatism	1.017
10	Eosin SumAverage mean	Cytoplasm Texture	Invasion of nuclei in the tumor	0.926

Table 4-6: Selected Features using mRMR and statistical test with Clinical interpretation

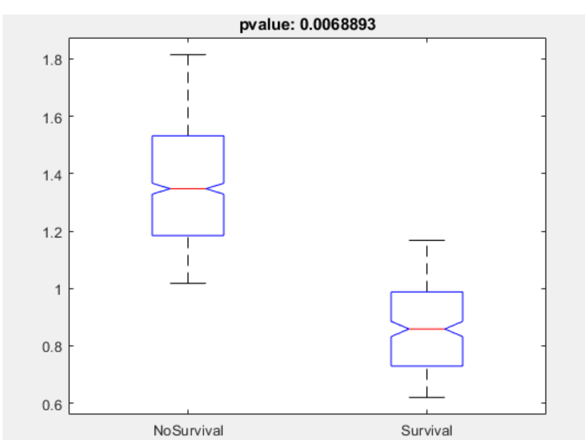

Figure 4-2: 1. Color / Texture

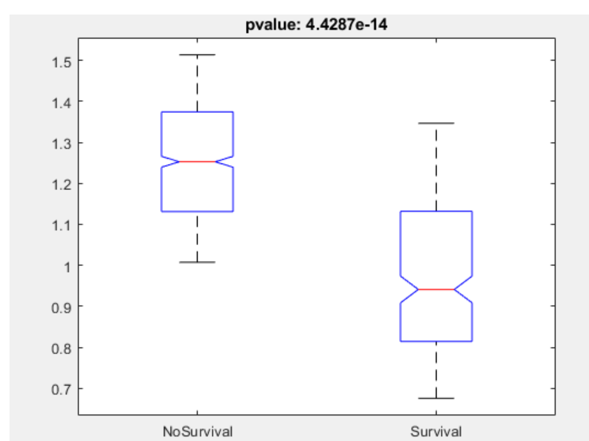

Figure 4-3: 2. Nuclei texture

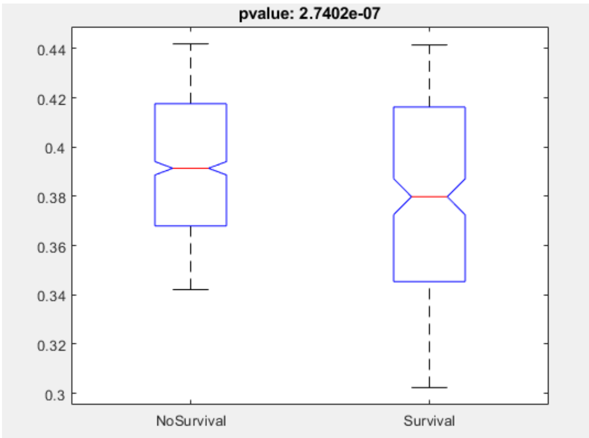

Figure 4-4: 3.Texture / Texture

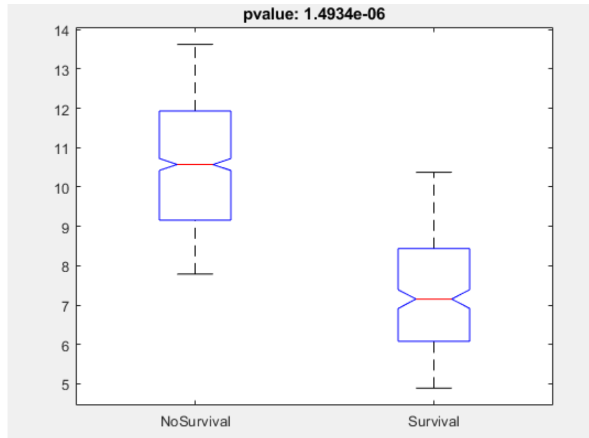

Figure 4-5: 4. Orientation / Shape cell

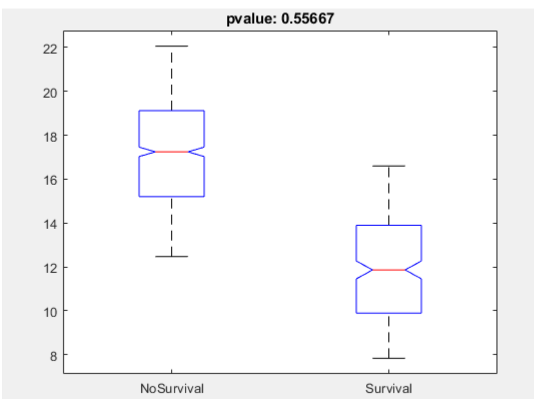

Figure 4-6: 5. Nuclei Color

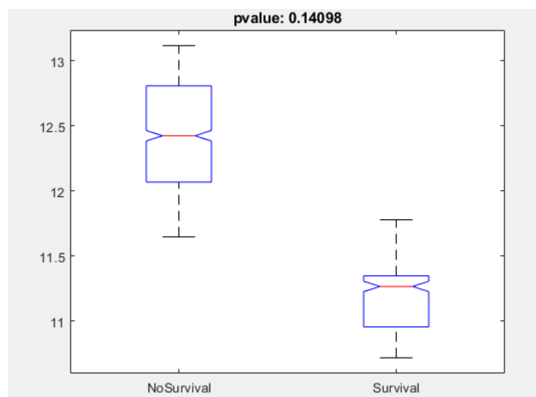

Figure 4-7: 6. Cytoplasm Texture

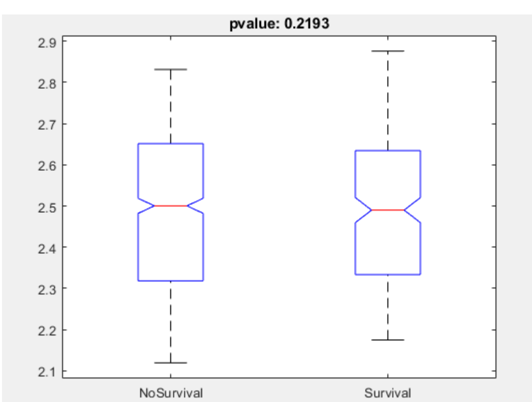

Figure 4-8: 7. Texture / Texture

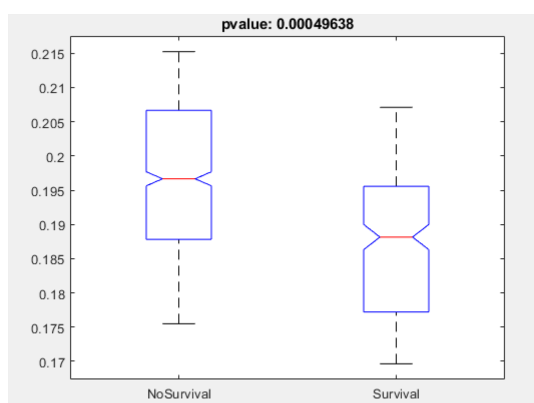

Figure 4-9: 8. Cytoplasm Texture

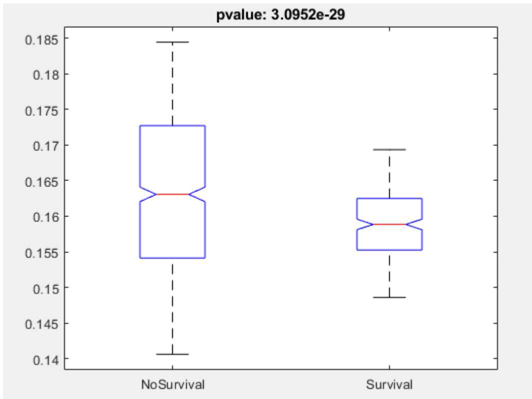

Figure 4-10: 9. Cytoplasm Texture

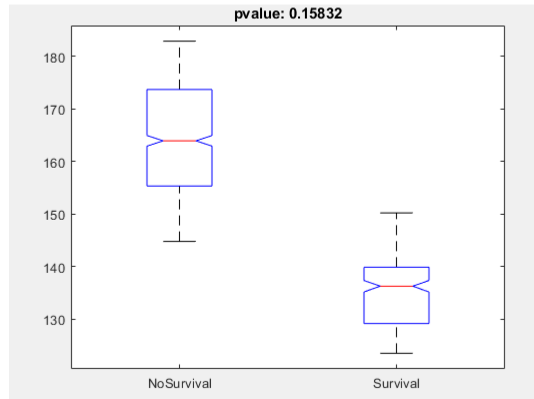

Figure 4-11: 10. Cytoplasm Texture

4.4. Conclusions

This work has proposed a complete framework to determine a set of features suitable for predicting survival time in GC patients. Yet 14 GC cases are a small sample, this work suggests they are discriminant enough as to separate aggressive cases from those with a more benign biological pattern. Interestingly, most relevant features highlight relations between morphometry and nuclei texture and between nuclei and glandular texture. Future extension of this work includes the use of an extensive database as The Carcinoma Genome Atlas [62].

5 Discussion and conclusion

This document presents first, two novel approaches that contribute to quantifying automatically glands in histopathological GC images, and second a study that statistic experiments show that exist relevant features that are correlated with the degree of the disease in spite of the small data set for evaluation. To discuss In the fourth chapter, we propose a complete approach to determine a set of major features suitable for predicting survival after a year in GC patients.
Exclusion criteria were applied to entire group of patients (35 cases), we do not have information of patient death-cause since a public page only reported the year-of-death, but this study found that the expert's diagnosis was report that most of these patients died in the same year. This probably related to the aggressiveness of the adenocarcinoma diagnosed. In general, it was not selected according to the diagnosis, it was used a reporting on the platform as deceased to 2019. Diagnoses were divided into two groups: survivors (6) and non-survivors (8). 14 GC cases are not enough for complete validation, nevertheless, our work suggests a set of discriminant features, that would be relevant in an extensive study. Unfortunately, we did not have a longitudinal study to make comparisons with the current diagnosis of each patient, but in further work, it will be taken into account to get a complement study.

Bibliografía

[1] ADACHI, Yosuke ; YASUDA, Kazuhiro ; INOMATA, Masafumi ; SATO, Koichi ; SHIRAISHI, Norio ; KITANO, Seigo: Pathology and prognosis of gastric carcinoma: well versus poorly differentiated type. En: *Cancer: Interdisciplinary International Journal of the American Cancer Society* 89 (2000), Nr. 7, p. 1418–1424

[2] ALFONSO, Sunny ; CORREDOR, Germán ; MONCAYO, Ricardo ; BARRERA, Cristian R. ; SANCHEZ, Angel Y. ; TORO, Paula ; ROMERO, Eduardo: A method to detect glands in histological gastric cancer images. En: *14th International Symposium on Medical Information Processing and Analysis* Vol. 10975 International Society for Optics and Photonics, 2018, p. 109750X

[3] ALVAREZ-ARELLANO, Lourdes ; CAMORLINGA-PONCE, Margarita ; MALDONADO-BERNAL, Carmen ; TORRES, Javier: Activation of human neutrophils with Helicobacter pylori and the role of Toll-like receptors 2 and 4 in the response. En: *FEMS Immunology & Medical Microbiology* 51 (2007), Nr. 3, p. 473–479

[4] ASSOCIATION, World M. [u. a.]: World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. En: *Bulletin of the World Health Organization* 79 (2001), Nr. 4, p. 373

[5] BARRERA, Cristian ; CORREDOR, Germán ; ALFONSO, Sunny ; MOSQUERA, Andrés ; ROMERO, Eduardo: An Automatic Segmentation of Gland Nuclei in Gastric Cancer Based on Local and Contextual Information. En: *Sipaim–Miccai Biomedical Workshop* Springer, 2018, p. 75–81

[6] BECKER, Carlos ; RIGAMONTI, Roberto ; LEPETIT, Vincent ; FUA, Pascal: Supervised feature learning for curvilinear structure segmentation. En: *International conference on medical image computing and computer-assisted intervention* Springer, 2013

[7] BERLTH, Felix ; BOLLSCHWEILER, Elfriede ; DREBBER, Uta ; HOELSCHER, Arnulf H. ; MOENIG, Stefan: Pathohistological classification systems in gastric cancer: diagnostic relevance and prognostic value. En: *World journal of gastroenterology: WJG* 20 (2014), Nr. 19, p. 5679

[8] BRAY, Freddie ; FERLAY, Jacques ; SOERJOMATARAM, Isabelle ; SIEGEL, Rebecca L. ; TORRE, Lindsey A. ; JEMAL, Ahmedin: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. En: *CA: a cancer journal for clinicians* 68 (2018), Nr. 6, p. 394–424

[9] CAMPAGNOLA, Paul J. ; LOEW, Leslie M.: Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. En: *Nature biotechnology* 21 (2003), Nr. 11, p. 1356

[10] CISŁO, Magdalena ; FILIP, Agata A. ; OFFERHAUS, George Johan A. ; CISEŁ, Bogumiła ; RAWICZ-PRUSZYŃSKI, Karol ; SKIERUCHA, Małgorzata ; POLKOWSKI, Wojciech P.: Distinct molecular subtypes of gastric cancer: from Laurén to molecular pathology. En: *Oncotarget* 9 (2018), Nr. 27, p. 19427

[11] MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL INSTITUTO NACIONAL DE CANCEROLOGÍA DE COLOMBIA, 2012-2021. *Plan Decenal para el Control de Cancer en Colombia, 2012-2021*. 2012

[12] MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL INSTITUTO NACIONAL DE CANCEROLOGÍA DE COLOMBIA, ESE]: *Plan Decenal para el Control de CÃ¡ncer en Colombia, 2012-2021*. 2012

[13] CORREA, P: Helicobacter pylori and gastric carcinogenesis. En: *The American journal of surgical pathology* 19 (1995), p. S37–43

[14] CORREA, Pelayo: Gastric cancer: overview. En: *Gastroenterology Clinics of North America* 42 (2013), Nr. 2, p. 211

[15] CORREA, Pelayo ; CUELLO, Carlos ; DUQUE, Edgar ; BURBANO, Luis C. ; GARCIA, Fernando T. ; BOLANOS, Oscar ; BROWN, Charles ; HAENSZEL, William: Gastric cancer in Colombia. III. Natural history of precursor lesions. En: *Journal of the National Cancer Institute* 57 (1976), Nr. 5, p. 1027–1035

[16] CORREA, Pelayo ; HAENSZEL, William ; CUELLO, Carlos ; TANNENBAUM, Steven ; ARCHER, Michael: A model for gastric cancer epidemiology. En: *The Lancet* 306 (1975), Nr. 7924, p. 58–60

[17] CORREA, Pelayo ; PIAZUELO, M B.: The gastric precancerous cascade. En: *Journal of digestive diseases* 13 (2012), Nr. 1, p. 2–9

[18] CORREA, Pelayo ; SCHNEIDER, Barbara G.: Etiology of gastric cancer: what is new? En: *Cancer Epidemiology and Prevention Biomarkers* 14 (2005), Nr. 8, p. 1865–1868

[19] COSATTO, Eric ; MILLER, Matt ; GRAF, Hans P. ; MEYER, John S.: Grading nuclear pleomorphism on histological micrographs. En: *2008 19th International Conference on Pattern Recognition* IEEE, 2008, p. 1–4

[20] CREW, Katherine D. ; NEUGUT, Alfred I.: Epidemiology of upper gastrointestinal malignancies. En: *Seminars in oncology* Vol. 31 Elsevier, 2004, p. 450–464

[21] DE VRIES, Annemarie C. ; VAN GRIEKEN, Nicole C. ; LOOMAN, Caspar W. ; CASPARIE, Mariël K ; DE VRIES, Esther ; MEIJER, Gerrit A. ; KUIPERS, Ernst J.: Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. En: *Gastroenterology* 134 (2008), Nr. 4, p. 945–952

[22] DOYLE, Scott ; HWANG, Mark ; SHAH, Kinsuk ; MADABHUSHI, Anant ; FELDMAN, Michael ; TOMASZEWESKI, John: Automated grading of prostate cancer using architectural and textural image features. En: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* IEEE, 2007, p. 1284–1287

[23] DUNN, Gavin P. ; OLD, Lloyd J. ; SCHREIBER, Robert D.: The immunobiology of cancer immunosurveillance and immunoediting. En: *Immunity* 21 (2004), Nr. 2, p. 137–148

[24] DUNN, Gavin P. ; OLD, Lloyd J. ; SCHREIBER, Robert D.: The three Es of cancer immunoediting. En: *Annu. Rev. Immunol.* 22 (2004), p. 329–360

[25] EKUNDINA, VO ; EZE, G: Common artifacts and remedies in histopathology (a review). En: *African Journal of Cellular Pathology* (2015), p. 1–7

[26] ERNST, Peter B. ; PEURA, David A. ; CROWE, Sheila E.: The translation of Helicobacter pylori basic research to patient care. En: *Gastroenterology* 130 (2006), Nr. 1, p. 188–206

[27] FAN, Xue-Gong ; YAKOOB, Javed ; FAN, XJ ; KEELING, PWN: Enhanced T-helper 2 lymphocyte responses: Immune mechanism ofHelicobacter pylori infection. En: *Irish journal of medical science* 165 (1996), Nr. 1, p. 37–39

[28] FERLAY, J ; SOERJOMATARAM, I ; ERVIK, Morten ; DIKSHIT, Rajesh ; ESER, Sultan ; MATHERS, C ; REBELO, M ; PARKIN, DM ; FORMAN, David ; BRAY, Freddie: GLOBOCAN 2012 v1. 0. En: *Cancer incidence and mortality worldwide: IARC CancerBase* 11 (2013)

[29] FICSOR, Levente ; VARGA, Viktor ; BERCZI, Lajos ; MIHELLER, Pal ; TAGSCHERER, Attila ; WU, Mark Li-cheng ; TULASSAY, Zsolt ; MOLNAR, Bela: Automated virtual microscopy of gastric biopsies. En: *Cytometry Part B: Clinical Cytometry: The Journal of the International Society for Analytical Cytology* 70 (2006), Nr. 6, p. 423–431

[30] FRIEDMAN, Jerome H.: Greedy function approximation: a gradient boosting machine. En: *Annals of statistics* (2001), p. 1189–1232

[31] GIRSHICK, Ross. *Fast r-cnn. ICCV 2015.* `https://arxiv.org/abs/1504.080838`. 2015

[32] GOBERT, Alain P. ; WILSON, Keith T.: The immune battle against Helicobacter pylori infection: NO offense. En: *Trends in microbiology* 24 (2016), Nr. 5, p. 366–376

[33] GOTINK, Annieke W. ; TEN KATE, Fiebo J. ; DOUKAS, Michael ; WIJNHOVEN, Bas P. ; BRUNO, Marco J. ; LOOIJENGA, Leendert H. ; KOCH, Arjun D. ; BIERMANN, Katharina: Do pathologists agree with each other on the histological assessment of pT1b oesophageal adenocarcinoma? En: *United European gastroenterology journal* 7 (2019), Nr. 2, p. 261–269

[34] GUNDUZ-DEMIR, Cigdem ; KANDEMIR, Melih ; TOSUN, Akif B. ; SOKMENSUER, Cenk: Automatic segmentation of colon glands using object-graphs. En: *Medical image analysis* 14 (2010), Nr. 1, p. 1–12

[35] HAMILTON., Booz A. *Kaggle: 2018 data science bowl.* `https://www.kaggle.com/c/data-science-bowl-2018`. 2018

[36] HAMILTON, Stanley R. ; AALTONEN, Lauri A. [u. a.]: *Pathology and genetics of tumours of the digestive system*. Vol. 48. IARC press Lyon:, 2000

[37] HAUB, Peter ; MECKEL, Tobias: A model based survey of colour deconvolution in diagnostic brightfield microscopy: Error estimation and spectral consideration. En: *Scientific reports* 5 (2015), p. 12096

[38] HE, Kaiming ; GKIOXARI, Georgia ; DOLLÁR, Piotr ; GIRSHICK, Ross: Mask r-cnn. En: *Proceedings of the IEEE international conference on computer vision*, 2017

[39] HOLMES, Rebecca S. ; VAUGHAN, Thomas L.: Epidemiology and pathogenesis of esophageal cancer. En: *Seminars in radiation oncology* Vol. 17 Elsevier, 2007, p. 2–9

[40] HU, Bing ; EL HAJJ, Nassim ; SITTLER, Scott ; LAMMERT, Nancy ; BARNES, Robert ; MELONI-EHRIG, Aurelia: Gastric cancer: Classification, histology and application of molecular pathology. En: *Journal of gastrointestinal oncology* 3 (2012), Nr. 3, p. 251

[41] JEMAL, Ahmedin ; BRAY, Freddie ; CENTER, Melissa M. ; FERLAY, Jacques ; WARD, Elizabeth ; FORMAN, David: Global cancer statistics. En: *CA: a cancer journal for clinicians* 61 (2011), Nr. 2, p. 69–90

[42] KAPADIA, Cyrus R.: Gastric atrophy, metaplasia, and dysplasia: a clinical perspective. En: *Journal of clinical gastroenterology* 36 (2003), Nr. 5, p. S29–S36

[43] KARIMI, Parisa ; ISLAMI, Farhad ; ANANDASABAPATHY, Sharmila ; FREEDMAN, Neal D. ; KAMANGAR, Farin: Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. En: *Cancer Epidemiology and Prevention Biomarkers* 23 (2014), Nr. 5, p. 700–713

[44] KUMAR, Vinay ; ABBAS, Abul K. ; FAUSTO, Nelson ; ASTER, Jon C.: *Robbins and Cotran pathologic basis of disease, professional edition e-book*. Elsevier health sciences, 2014

[45] LAUREN, Pekka: The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma: an attempt at a histo-clinical classification. En: *Acta Pathologica Microbiologica Scandinavica* 64 (1965), Nr. 1, p. 31–49

[46] MACENKO, Marc ; NIETHAMMER, Marc ; MARRON, James S. ; BORLAND, David ; WOOSLEY, John T. ; GUAN, Xiaojun ; SCHMITT, Charles ; THOMAS, Nancy E.: A method for normalizing histology slides for quantitative analysis. En: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* IEEE, 2009, p. 1107–1110

[47] MAI, UE ; PEREZ-PEREZ, GI ; WAHL, LM ; WAHL, SM ; BLASER, MJ ; SMITH, PD: Soluble surface proteins from Helicobacter pylori activate monocytes/macrophages by lipopolysaccharide-independent mechanism. En: *The Journal of clinical investigation* 87 (1991), Nr. 3, p. 894–900

[48] MARQUÉS-LESPIER, Juan M. ; GONZÁLEZ-PONS, María ; CRUZ-CORREA, Marcia: Current perspectives on gastric cancer. En: *Gastroenterology Clinics* 45 (2016), Nr. 3, p. 413–428

[49] MARTIN, IG ; DIXON, MF ; SUE-LING, H ; AXON, AT ; JOHNSTON, D: Goseki histological grading of gastric cancer is an important predictor of outcome. En: *Gut* 35 (1994), Nr. 6, p. 758–763

[50] MORALES ÁLVAREZ, Anamaría [u. a.]: *Inmuno-monitoreo del componente de células presentadoras de antígeno (APC) y células T en distintos estadios del desarrollo de cáncer gástrico de tipo intestinal*, Universidad Nacional de Colombia-Sede Bogotá, Tesis de Grado

[51] 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. En: *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* IEEE, 2008, p. 284–287

[52] OH, SE ; CHOI, MG ; SEO, SW ; SOHN, TS ; BAE, JM ; KIM, S: Prediction of overall survival and novel classification of patients with gastric cancer using the survival recurrent network. En: *European Journal of Surgical Oncology* 45 (2019), Nr. 2, p. e79–e80

[53] PARKIN, D M.: Global cancer statistics in the year 2000. En: *The lancet oncology* 2 (2001), Nr. 9, p. 533–543

[54] PARKIN, Donald M.: International variation. En: *Oncogene* 23 (2004), Nr. 38, p. 6329

[55] PENG, Hanchuan ; LONG, Fuhui ; DING, Chris: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. En: *IEEE Transactions on Pattern Analysis & Machine Intelligence* (2005), Nr. 8, p. 1226–1238

[56] RABINOVICH, Gabriel A. ; GABRILOVICH, Dmitry ; SOTOMAYOR, Eduardo M.: Immunosuppressive strategies that are mediated by tumor cells. En: *Annu. Rev. Immunol.* 25 (2007), p. 267–296

[57] ROFFO, Giorgio ; MELZI, Simone ; CASTELLANI, Umberto ; VINCIARELLI, Alessandro: Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach. En: *2017 IEEE International Conference on Computer Vision (ICCV)* IEEE, 2017, p. 1407–1415

[58] SEMPER, Raphaela P. ; MEJÍAS-LUQUE, Raquel ; GROSS, Christina ; ANDERL, Florian ; MÜLLER, Anne ; VIETH, Michael ; BUSCH, Dirk H. ; DA COSTA, Clarissa P. ; RULAND, Jürgen ; GROSS, Olaf [u. a.]: Helicobacter pylori–induced IL-1β secretion in innate immune cells is regulated by the NLRP3 Inflammasome and requires the cag Pathogenicity Island. En: *The Journal of Immunology* 193 (2014), Nr. 7, p. 3566–3576

[59] SPARKS, Rachel ; MADABHUSHI, Anant: Explicit shape descriptors: Novel morphologic features for histopathology classification. En: *Medical image analysis* 17 (2013), Nr. 8, p. 997–1009

[60] SUN, Gaofeng ; CHENG, Chao ; LI, Xiao ; WANG, Tao ; YANG, Jian ; LI, Danni: Metabolic tumor burden on postsurgical PET/CT predicts survival of patients with gastric cancer. En: *Cancer Imaging* 19 (2019), Nr. 1, p. 18

[61] TEKESIN, Kemal ; GUNES, Mehmet E. ; TURAL, Deniz ; AKAR, Emre ; ZIRTILOGLU, Alisan ; KARACA, Mustafa ; SELCUKBIRICIK, Fatih ; BAYRAK, Savas ; OZET, Ahmet: Clinicopathological characteristics, prognosis and survival outcome of gastric cancer in young patients: A large cohort retrospective study. En: *future* 1 (2019), p. 3

[62] TOMCZAK, Katarzyna ; CZERWIŃSKA, Patrycja ; WIZNEROWICZ, Maciej: The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. En: *Contemporary oncology* 19 (2015), Nr. 1A, p. A68

[63] TORRE, Lindsey A. ; BRAY, Freddie ; SIEGEL, Rebecca L. ; FERLAY, Jacques ; LORTET-TIEULENT, Joannie ; JEMAL, Ahmedin: Global cancer statistics, 2012. En: *CA: a cancer journal for clinicians* 65 (2015), Nr. 2, p. 87–108

[64] VAN CUTSEM, Eric ; SAGAERT, Xavier ; TOPAL, Baki ; HAUSTERMANS, Karin ; PRENEN, Hans: Gastric cancer. En: *The Lancet* 388 (2016), Nr. 10060, p. 2654–2664

[65] VETA, Mitko ; VAN DIEST, Paul J. ; KORNEGOOR, Robert ; HUISMAN, André ; VIERGEVER, Max A. ; PLUIM, Josien P.: Automatic nuclei segmentation in H&E stained breast cancer histopathology images. En: *PloS one* 8 (2013), Nr. 7, p. e70221

[66] WANG, Fei ; MENG, Wenbo ; WANG, Bingyuan ; QIAO, Liang: Helicobacter pylori-induced gastric inflammation and gastric cancer. En: *Cancer letters* 345 (2014), Nr. 2, p. 196–202

[67] WARREN, JR: Marshall. 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. En: *Lancet* , p. 1273–1275

[68] WILLIAMS, C ; POLOM, K ; ADAMCZYK, B ; AFSHAR, M ; D'IGNAZIO, A ; KAMALI-MOGHADDAM, M ; KARLSSON, NG ; GUERGOVA-KURAS, M ; LISACEK, F ; MARRELLI, D [u. a.]: Machine learning methodology applied to characterize subgroups of gastric cancer patients using an integrated large biomarker dataset. En: *European Journal of Surgical Oncology* 45 (2019), Nr. 2, p. e79

[69] YOSHIDA, Hiroshi ; SHIMAZU, Taichi ; KIYUNA, Tomoharu ; MARUGAME, Atsushi ; YAMASHITA, Yoshiko ; COSATTO, Eric ; TANIGUCHI, Hirokazu ; SEKINE, Shigeki ; OCHIAI, Atsushi: Automated histological classification of whole-slide images of gastric biopsy specimens. En: *Gastric Cancer* 21 (2018), Nr. 2, p. 249–257