SEGMENTACIÓN DE PATRONES ASOCIADOS AL CÁNCER DE PRÓSTATA, SIGUIENDO LA ESCALA DE GLEASON Y UTILIZANDO REPRESENTACIONES PROFUNDAS

ANDRÉS FELIPE GÓMEZ ORTIZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2021

SEGMENTACIÓN DE PATRONES ASOCIADOS AL CÁNCER DE PRÓSTATA, SIGUIENDO LA ESCALA DE GLEASON Y UTILIZANDO REPRESENTACIONES PROFUNDAS

ANDRÉS FELIPE GÓMEZ ORTIZ

Una tesis presentada en cumplimiento de los requisitos para el grado de:
Ingeniero de Sistemas e Informática

Director:
Fabio Martínez Carrillo
Ph.D en Ingeniería de Sistemas y Computación

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2021

AGRADECIMIENTOS

El autor expresa su agradecimiento:

Primeramente a Dios, por haberme dado la fortaleza para seguir adelante, a mi padre Edgar Gómez, quien ha sido mi apoyo más grande durante mi educación universitaria, quien me ha brindado consejos y me ha guiado a lo largo de mi vida para poder cumplir mis sueños y ser una gran persona.

A mi madre Stella Ortiz, por estar siempre presente a lo largo de mi vida, por apoyarme en todo lo que me he propuesto y por impulsarme a ser mejor cada día.

A mis hermanos Walter, Astrid y Sandra, quienes de alguna manera me han demostrado su apoyo incondicional y han estado presentes para todo lo que necesite.

A mi familia, por estar pendientes de mi y mostrarme su apoyo.

A mi director de proyecto de grado, Fabio Martínez, quien es un ejemplo a seguir, quien me impulsó a seguir adelante y a trabajar fuerte durante la realización del mismo, quien me forjó en mi camino de la investigación, y por enseñarme que un gran esfuerzo al final conlleva a grandes frutos.

A los Ingenieros Fabián León y Miguel Plazas, que han sido mis tutores, quienes son unas grandes personas y me brindado su apoyo en cualquier duda que tuve en el presente proyecto, guiándome y haciéndome crecer en el camino de la investigación.

A mis amigos Christian, Juan y Andrea, quienes han sido un gran apoyo a lo largo de mi etapa

universitaria y han estado presentes a pesar de todo, y a una persona muy especial, Karen, quien ha sido mi ayuda fundamental a lo largo de este proceso, quien siempre con motivación, en los mejores y peores momentos, estuvo brindándome su apoyo a pesar de cualquier adversidad, y quien siempre con palabras esperanzadoras me ayudó a seguir adelante. Estoy agradecido por haberlos conocido y por poder compartir este logro junto a ellos.

A mis compañeros del grupo de investigación $BIVL^2ab$, quienes a partir de su conocimiento he logrado un gran aprendizaje que me ha hecho desarrollarme en los ámbitos personal y profesional.

Índice general

Índice de figuras

Pág

Índice de cuadros

Pág

RESUMEN

TÍTULO: SEGMENTACIÓN DE PATRONES ASOCIADOS AL CÁNCER DE PRÓSTATA, SIGUIENDO LA ESCALA DE GLEASON Y UTILIZANDO REPRESENTACIONES PROFUNDAS. [*]

AUTOR: ANDRÉS FELIPE GÓMEZ ORTIZ [**]

PALABRAS CLAVE: Segmentación, representaciones profundas, escala de Gleason, imágenes histopatológicas, cáncer de próstata.

DESCRIPCIÓN: El análisis histológico es la principal herramienta para diagnosticar y cuantificar la agresividad del cáncer de próstata. El sistema de puntuación de Gleason es el sistema más utilizado para cuantificar la agresividad de la enfermedad sobre histologías. Este sistema permite estratificar regionalmente los patrones anormales en las placas histológicas, dando pautas para la puntuación y grado de la enfermedad. A pesar de ello, estudios recientes han mostrado una variabilidad persistente en el diagnóstico de la enfermedad, reportando valores moderados de concordancia de 0.55, según el valor kappa.

Este trabajo introduce un enfoque de segmentación y estratificación de regiones de acuerdo con las segmentaciones realizadas siguiendo el puntaje de Gleason. En un primer nivel, una red de aprendizaje profundo regional es entrenada con anotaciones completas, sobre imágenes histopatológicas, realizadas por un experto patólogo. Esta arquitectura permite definir delineaciones regionales, siendo efectivo en localizaciones con estructuras generales. En un segundo nivel de representación, se entrenó un modelo únicamente con anotaciones superpuestas del primer esquema, y que constituyen regiones con dificultad de clasificación. Finalmente, en un tercer nivel de representación, que permite una descripción más granular de las regiones, se entrenó una tercera red con las regiones resultantes de las activaciones de la representación del primer nivel. La segmentación final resulta entonces de la superposición de los tres niveles de representación. La estrategia propuesta se validó en un conjunto público con 886 imágenes correspondientes a microarreglos de tejidos histológicos con anotaciones de grados de Gleason: benigno, 3, 4 y 5. Las segmentaciones generadas lograron en promedio un AUPRC (área bajo la curva de precisión-sensibilidad *precision-recall*) de 0.8 con respecto al diagnóstico de un primer patólogo, y de 0.76 con respecto al diagnóstico de un segundo patólogo.

---

[*] Trabajo de investigación

[**] Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez Carrillo, Ph.D.

ABSTRACT

TITLE: SEGMENTATION OF PATTERNS ASSOCIATED WITH PROSTATE CANCER, FOLLOWING THE GLEASON SCALE AND USING DEEP REPRESENTATIONS. *

AUTHOR: ANDRÉS FELIPE GÓMEZ ORTIZ **

KEYWORDS: Segmentation, deep representations, Gleason scale, histopathologic images, prostate cancer.

DESCRIPTION: Histological analysis is the main tool for diagnosing and quantifying the aggressiveness of prostate cancer. The Gleason scoring system is the most widely used system to quantify the aggressiveness of the disease on histologies. This system allows regional stratification of abnormal patterns in histologic plaques, providing guidelines for scoring and grading of disease. Despite this, recent studies have shown persistent variability in disease diagnosis, reporting moderate concordance values of 0.55, according to the kappa value.

This work introduces an approach of segmentation and stratification of regions according to the segmentations performed following the Gleason score. At a first level, a regional deep learning network is trained with complete annotations, on histopathological images, performed by an expert pathologist. This architecture allows defining regional delineations, being effective in locations with general structures. In a second level of representation, a model was trained only with overlapping annotations of the first scheme, which constitute regions with difficult classification. Finally, in a third level of representation, which allows a more granular description of the regions, a third network was trained with the regions resulting from the activations of the first level representation. The final segmentation then results from the superposition of the three levels of representation. The proposed strategy was validated on a public set with 886 images corresponding to histological tissue microarrays with Gleason grade annotations: benign, 3, 4 and 5. The generated segmentations achieved on average an AUPRC (area under the precision-recall curve) of 0.8 with respect to the diagnosis of a first pathologist, and 0.76 with respect to the diagnosis of a second pathologist.

---

* Research work

** Faculty of Physical-Mechanical Engineering. School of Systems and Computer Engineering. Advisor: Fabio Martínez Carrillo

INTRODUCCIÓN

El cáncer de próstata es el cuarto cáncer más frecuente en el mundo, con mas de un millón doscientos mil nuevos casos y más de trescientas mil muertes cada año [1]. Existen diversas herramientas diagnósticas como la resonancia magnética, ecografía transrectal, pero con reportadas limitaciones de especificidad en la tarea de detección, caracterización y prógnosis de la enfermedad [2]. Es por ello que hoy en día, las biopsias constituyen el principal método para cuantificar la agresividad de la enfermedad, mediante un análisis histológico de imágenes microscópicas. Estas imágenes son obtenidas mediante la tinción de muestras con hematoxilina y eosina (H&E), que permiten destacar y caracterizar la distribución arquitectural de las células y la geometría atípica de estructuras glandulares [3].

El sistema de puntuación de Gleason es el principal método de apoyo para los patólogos en cuanto a la cuantificación, la estandarización del diagnóstico y la descripción de patrones característicos relacionados con la evolución de la enfermedad. Para ello, este sistema se basa principalmente en el análisis de estructuras glandulares, definiéndo su estándar en dos escalas de puntuación. La primera escala esta dedicada a caracterizar variaciones atípicas en patrones visuales con puntuaciones entre uno y cinco. En una segunda escala de Gleason se determina el grado de afectación y progresión de la enfermedad, según la suma de los valores predominantes de la primera escala. Esta segunda escala de diagnóstico está acotada entre valores de seis y diez. Específicamente, en esta segunda escala se obtiene una valoración de la muestra mediante la suma entre los dos grados más comunes en la muestra. Sin embargo, el procedimiento de este

---

[1] Bray, F. y col. "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(6) (2018), págs. 394-424.

[2] Enrique Bley y Andrés Silva. "Diagnóstico precoz del cáncer de próstata". En: *Revista médica clínica Las Condes* 22.4 (2011), págs. 453-458.

[3] Ana Isabel Ruiz y col. "Actualización sobre cáncer de próstata". En: *Correo Científico Médico* 21.3 (2017).

sistema es tedioso, estimándose que para un experto puede tomar hasta tres días en el etiquetado de 6 a 15 muestras de una sola biopsia [4]. Además, en la rutina clínica este procedimiento es totalmente dependiente de la interpretación visual del experto, lo que introduce una gran variabilidad en el diagnóstico. Diversos estudios han reportado valores bajos y moderados de concordancia, que van desde puntajes promedios de 0.55 (sobre 30 muestras y tres patólogos) hasta valores bajos de 0.33 (sobre 20 muestras y 24 patólogos)[5][6][7].

En la literatura han emergido diferentes propuestas computacionales y sistemas de apoyo que buscan mitigar esta variabilidad y subjetividad en el diagnóstico. Por ejemplo, en [8] se utilizaron 102 características texturales y morfológicas para clasificar epitelio benigno, estroma benigno, grado 3 y grado 4 de Gleason. Sin embargo, estos enfoques no suelen generalizar completamente la enfermedad al limitarse sólo a cierto número de características predefinidas. Durante los últimos años, los enfoques basados en aprendizaje profundo han mostrado grandes resultados en diferentes campos de aplicación, incluyendo la histología. Por ejemplo, en [9] se entrenó una red neuronal convolucional (CNN) en un conjunto de más de 800 imágenes para la clasificación de parches benignos, Gleason 3, 4 y 5 obteniendo una exactitud del 70 %. Otros enfoques en redes

---

[4] American Cancer Society. *Pruebas para diagnosticar y determinar la etapa del cáncer de próstata*. En: cancer.org. 2019.

[5] D F R Griffiths y col. "A study of Gleason score interpretation in different groups of UK pathologists; techniques for improving reproducibility". En: *Histopathology* 48.6 (2006), págs. 655-662.

[6] Comisión de Salud Pública. "Evaluación del Sistema Gleason". En: *Urología Colombiana* IX.1 ().

[7] M McLean y col. "Interobserver variation in prostate cancer gleason scoring: are there implications for the design of clinical trials and treatment strategies?" En: *Clinical oncology* 9.4 (1997), págs. 222-225.

[8] S. Doyle y col. "Automated Grading of Prostate Cancer using Architectural and Textural Image Features". En: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2007), págs. 1284-1287.

[9] Eirini Arvaniti y col. "Automated Gleason grading of prostate cancer tissue microarrays via deep learning". En: *Scientific reports* (2018).

neuronales incluyen técnicas de creación de parches de imágenes [10], extracción de características por medio de filtrado de imágenes[11], detección de regiones de interés [12] y segmentación de glándulas [13]. Estas estrategias se centran en una caracterización local de la enfermedad, dividiendo las imágenes histológicas en parches, lo cual va en contravía o resulta insuficiente, según lo definido en la escala de Gleason.

Este trabajo introduce una estrategia para segmentar y estratificar patrones en imágenes histológicas según la escala de Gleason. En este sentido, los puntajes de Gleason son asociados a patrones espaciales particulares, que pueden ser segmentados como regiones coherentes en una imagen histológica. Se implementó una estrategia multinivel utilizando representaciones con el método Mask R-CNN, un esquema de segmentación supervisado basado en redes neuronales convolucionales (CNN). En un primer nivel de procesamiento, se toman todos los segmentos delineados por patólogos, asociados a diferentes grados de Gleason, y se ajusta una representación profunda de tipo Mask R-CNN. Esta arquitectura entrenada genera delineaciones regionales, siguiendo un grado de Gleason y con una probabilidad asociada de predicción. Sin embargo, para ciertas regiones se pueden obtener multiples segmentaciones para una misma región debido a la alta variabilidad del problema. Entonces, se definió una segunda representación de tipo Mask R-CNN dedicada únicamente a definir segmentaciones en regiones desafiantes, con múltiples candidatos de grado para una misma localización. Finalmente, en un tercer nivel de representa-

---

[10] Fabian León, Miguel Plazas y Fabio Martínez. "An inception deep architecture to differentiate close-related Gleason prostate cancer scores". En: (2019).

[11] Scott Doyle y col. "A Boosting Cascade for Automated Detection of Prostate Cancer from Digitized Histology". En: *Dept. of Biomedical Engineering, Rutgers Univ., Piscataway, NJ 08854, USA, Dept. of Surgical Pathology, Univ. of Pennsylvania, Philadelphia, PA 19104, USA* IX.1 ().

[12] Elena Payá, Valery Naranjo y Jose García. "Diseño y desarrollo de un sistema automático de segmentación de glándulas histológicas para identificar el cáncer de próstata en una etapa inicial". En: *Universidad Politécnica de Valencia* (2019), pág. 5.

[13] José Gabriel García, Valery Naranjo y Adrián Colomer. "Diseño y desarrollo de un sistema automático de clasificación de estructuras glandulares en imágenes histológicas de próstata". En: *Universidad Politécnica de Valencia* (2018).

ción, se entrenó una representación con un enfoque regional y granular, tomando como regiones de entrada segmentos con mayor atención, en las capas de atención, de la representación del primer nivel. La superposición de los tres niveles de representación conlleva a la segmentación final. Esta estrategia puede soportar los diagnósticos en la escala de Gleason, así como también permite proponer marcaciones iniciales a los patólogos para agilizar su tarea de análisis.

1. MARCO TEÓRICO Y TRABAJOS PREVIOS

1.1. Sistema de puntuación de Gleason

La detección de anormalidades en la próstata y el diagnóstico precoz del cáncer, se realiza típicamente mediante técnicas clínicas como tacto rectal o muestras de sangre para detección de antígeno prostático específico [14]. También, existen alternativas clínicas adicionales que se apoyan en la caracterización de la enfermedad, usando imágenes por resonancia magnética o ecografía transrectal. A pesar de que estos procedimientos son las alternativas primarias de diagnóstico, existen limitaciones relacionadas con un bajo valor predictivo positivo y poca especificidad [3].
El análisis histológico es la alternativa clínica más confiable para el diagnóstico y caracterización del cáncer. En este análisis se valora el tejido prostático para determinar y caracterizar la existencia de células tumorales. Estas muestras son recuperadas mediante procedimientos de biopsia donde se extrae un cilindro de tejido (normalmente de 1cm de longitud y de 2mm de ancho) que posteriormente se utiliza en un análisis histopatológico [15]. En el laboratorio, estas muestras prostáticas son artificialmente coloreadas para resaltar estructuras, características morfológicas y componentes celulares como el citoplasma (la tinción es hematoxilina y eosina H&E) [16]. Durante el análisis histopatológico, el experto observa estas muestras con magnificaciones típicas de 20× y 40×, tomando algunas veces hasta 96 horas para realizar un análisis completo de la biopsia [17].
El sistema de puntuación de Gleason es el sistema más usado para la cuantificación de la

---

[14] Christian Ramos, Juan Fullá y Alejandro Mercado. "Detección precoz de cáncer de próstata: Controversias y recomendaciones actuales". En: *Revista médica clínica Las Condes* 29.2 (2018), págs. 128-135.

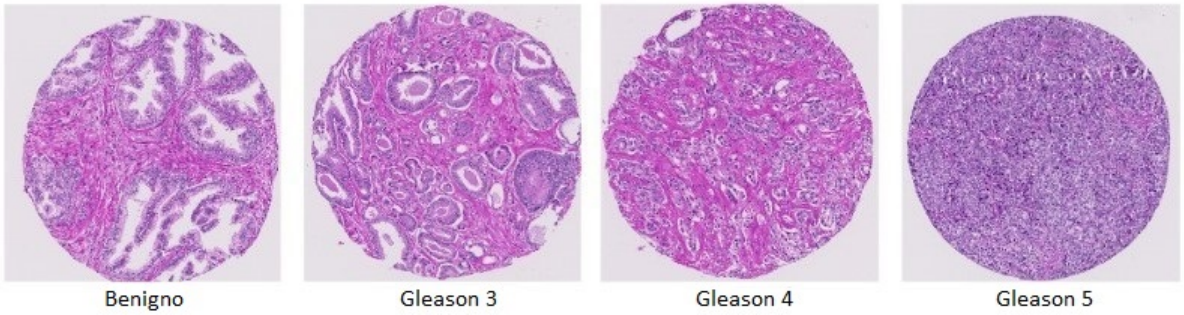
[15] American Cancer Society. *Cómo entender su informe de patología: cáncer de próstata.* En: cancer.org. 2017.

[16] Instituto Nacional del Cáncer. *Tinción con hematoxilina y eosina.*

[17] Aziza Nassar. "Biopsia: 5 cosas que todo paciente debe saber". En: *Cancer.Net* (2017).

agresividad del cáncer de próstata [18]. Este sistema se basa principalmente en características glandulares, asignando un grado dependiendo de la anormalidad presente en el tejido [4].

Figura 1. Muestras de microarray de tejido (TMA) marcadas por el primer patólogo de izquierda a derecha como: Benigno, Gleason 3, Gleason 4 y Gleason 5.



La puntuación de Gleason se divide en dos escalas de puntuación. En la primera escala, como se puede observar en la Figura 1, se encuentran los puntajes (o grados) entre uno y cinco (1-5), que permiten cuantificar regiones locales. En este caso, los grados uno y dos están asociados a las células prostáticas definidas como poco agresivas, observadas con glándulas bien formadas y con patrones similares al tejido sano. En este sentido, estos patrones celulares tienen una repercusión clínica mínima y no suelen ser reportados como cáncer. El grado tres sigue indicando un patrón similar al tejido normal, pero se aprecian roturas de glándulas e infiltración celular. El grado 4 se suele caracterizar por presentar por glándulas pobremente formadas y patrones cribiformes. Finalmente, en el grado 5, el tejido carece de glándulas formadas, no tiene estructura propia, las células prostáticas son muy diferentes a las normales y posee la mayor malignidad [12].
El segundo nivel de graduación en la escala de Gleason, de seis a diez (6-10), cuantifica regiones globales proporcionando un diagnóstico global de toda la muestra prostática. Esta escala se cuantifica como la suma de los dos grados más predominantes en la muestra. Por ejemplo, para una muestra de tejido en donde el grado predominante en las células es 3, seguido del grado

---

[18] Pamela Bolaños Morera y Carolina Chacón Araya. "Escala patológica de Gleason para el cáncer de prostata y sus modificaciones". En: *Medicina Legal de Costa Rica* 34.1 (2017).

4, se le asigna una puntuación de Gleason 7 (3+4). El orden de los factores es importante, ya que el primero indica el grado predominante de la imagen, y el segundo el siguiente grado mayoritario. Este segundo nivel de graduación determina la severidad del cancer y permite una aproximación a la cuantificación de la enfermedad. En función de ello, la severidad del cáncer varía y por consiguiente, el tratamiento también [15].

A pesar de que este sistema se establece como la herramienta primordial para la selección del tratamiento del cáncer de próstata, sigue siendo un procedimiento tedioso, demandando para un experto patólogo hasta tres días en el etiquetado de seis a quince muestras de una sola biopsia [4]. Este procedimiento además introduce subjetividad, dependiendo la graduación de la interpretación visual del experto, lo que introduce una gran variabilidad en el diagnóstico de la enfermedad. Esta subjetividad y variabilidad en el diagnóstico ha sido estudiada y reportada en diferentes estudios del estado del arte. Por ejemplo, en [6], se enviaron 30 muestras a un grupo de tres patólogos para su estudio, y se obtuvo una concordancia máxima (según el valor kappa ponderado) de 0.5517 entre grados contiguos. En [5] se informó de que, para veinte diapositivas en rangos de los grupos de puntuación de Gleason 2-4, 5-6, 7 y 8-10, enviados a 24 patólogos, la concordancia entre observadores (kappa ponderado) fue de 0.33. Asimismo, en [7], el estadístico kappa también se utilizó para evaluar el nivel de acuerdo entre tres patólogos, en 71 diapositivas histológicas, siendo el valor máximo de concordancia 0.29 para las puntuaciones de Gleason de 2-10. Los estudios mencionados anteriormente, muestran la variabilidad de este diagnóstico y destacan la necesidad de mejorar la reproducibilidad del observador, ya que este estándar constituye el principal mecanismo para la selección del tratamiento de la enfermedad y ayuda a predecir el pronóstico de un paciente, es decir, la probabilidad de recuperación.

1.2. Métodos de representación y soporte en el análisis histológico

Los métodos de soporte al diagnóstico han sido ampliamente beneficiados con los avances de escáneres digitales, lo cuales permiten ampliar, codificar y analizar histopatologías digitales. Sumado a esto, los recientes desarrollos en análisis de imágenes y métodos de aprendizaje de

máquina han permitido determinar patrones para valorar automáticamente anormalidades presentes en las placas, sirviendo así para reducir la variabilidad en el diagnóstico del cáncer de próstata [19]. En la literatura existen diferentes propuestas computacionales relacionadas con la cuantificación y la caracterización del cáncer de próstata, mitigando la variabilidad y subjetividad en el diagnóstico. Estas metodologías abarcan diversos enfoques tales como extracción de características por medio de filtrado de imágenes, creación de parches de imágenes y segmentación de glándulas. En las siguientes subsecciones se presentará un resumen de estas metodologías para poner en contexto la siguiente propuesta.

Ingeniería de características para la clasificación de grados de Gleason: Una metodología clásica para el reconocimiento y análisis de patrones visuales es el diseño y codificación de características típicas, asociadas a cambios visuales y geométricos en las histologías. Por ejemplo, Doyle et al. [11] proponen un CAD totalmente automatizado para la clasificación de tejido benigno y maligno utilizando una descomposición piramidal de las histologías, permitiendo desarrollar un análisis en diferentes escalas de representación. En este trabajo, se extraen cerca de 600 características de textura en cada descomposición para su posterior clasificación, utilizando una clasificación en cascada. En [20] se propone un esquema de segmentación de glándulas aplicando un filtro de varianza que calcula características asociadas al tamaño y la forma de las glándulas para generar un índice, proporcional a la malignidad del cáncer. En [8], fueron caracterizadas 75 imágenes histopatológicas con puntuaciones de Gleason 3 y 4, utilizando una triangulación de Delaunay. De este modo, permite localizar núcleos individuales para la extracción de 102 características morfológicas y de textura, que se utilizaron para la posterior clasificación con una SVM (Support Vector Machine). Estos enfoques, sin embargo, no suelen

---

[19] Metin N Gurcan y col. "Histopathological image analysis: A review". En: *IEEE reviews in biomedical engineering* 2 (2009), págs. 147-171.

[20] R. Farjam y col. "An Image Analysis Approach for Automatic Malignancy Determination of Prostate Pathological Images". En: *Cytometry. Part B, Clinical cytometry* 72 (2007), págs. 227-40.

generalizar completamente la enfermedad al limitarse a un cierto número de características predefinidas, y dependen de las normalizaciones de color, el tamaño de las muestras y el aumento, entre otras dependencias.

Enfoques basados en representaciones profundas: En los últimos años, los enfoques de aprendizaje profundo han dado grandes resultados en diferentes campos de aplicación, incluyendo la histología. Por ejemplo, Arvaniti et al. [9] propusieron una red neuronal convolucional, tomando pesos preentrenados de un conjunto de imágenes generales. Estas se utilizaron para la clasificación de tejido benigno y tejido con grados Gleason 3, 4 y 5, utilizando parches de tamaño 750x750. En [10], se implementó una red InceptionV3, que hace uso de la factorización del kernel y las conexiones residuales para profundizar, sin sufrir desvanecimiento de gradiente. Esta red se utilizó para clasificar los grados de Gleason 3 y 4 en parches histológicos. Sin embargo, estos enfoques basados en parches se basan únicamente en una caracterización local, perdiendo componentes estructurales del tejido que podrían ser definitivos para la graduación apropiada de una muestra.

Cuantificación y segmentación regional en placas histológicas: Otros enfoques se han dedicado recientemente a la segmentación de glándulas cancerígenas en imágenes histológicas. Por ejemplo, Bulten et al. [21] proponen un método de segmentación de glándulas en imágenes histológicas utilizando tres redes convolucionales para la detección de tumor, detección de tejido no epitelial y, finalmente, para la segmentación de las glándulas prostáticas. Posteriormente, se clasifican cada una de las glándulas para calcular el porcentaje de tumor en cada uno de los grados. Por otra parte, en el trabajo de Nathan et al. [22] se implementaron cuatro arquitecturas

---

[21] Wouter Bulten y col. "Automated gleason grading of prostate biopsies using deep learning". En: *arXiv preprint arXiv:1907.07980* (2019).

[22] Nathan Ing y col. "Semantic segmentation for prostate cancer grading by convolutional neural networks". En: *ResearchGate* (2018).

de redes neuronales convolucionales: FCN-8s, dos variantes de SegNet y U-Net multiescala, para la segmentación semántica de tumores de alto y bajo grado en glándulas, según la escala de Gleason. Este trabajo en sus imágenes presenta anotaciones a nivel de glándulas de tejido benigno y grados de Gleason 3, 4 y 5, por lo que en sus resultados igualmente se generan anotaciones o segmentaciones glandulares. Estos análisis podrían contribuir en la cuantificación de la enfermedad al caracterizar glándulas en distintos grados de Gleason.

1.3. Métodos de detección de objetos utilizando aprendizaje profundo

En el aprendizaje profundo, las estrategias están compuestas por nodos organizados jerárquicamente que permiten representar problemas complejos de aprendizaje [23]. Particularmente, las redes neuronales convolucionales (CNN) aprenden múltiples filtros que operan localmente para extraer patrones, aprender características y representar conceptos codificados, normalmente, en imágenes [24].

Desde el punto de vista del aprendizaje profundo, la segmentación densa consiste en asignar un objeto de clase a cada uno de los píxeles de la imagen. Por tanto, la salida es una imagen coloreada con las diferentes instancias clasificadas a nivel de píxel. A continuación, se realiza una descripción de distintos métodos de detección de objetos basados en representaciones profundas.

R-CNN (Redes neuronales convolucionales basadas en regiones): Esta estrategia fue pionera en el uso de representaciones CNN para detectar objetos en imágenes [25]. Inicialmente, en la estrategia se utiliza un método de búsqueda selectiva de regiones de interés, usando agrupamiento de píxeles y decisiones basadas en grafos. Este algoritmo utiliza ventanas desli-

---

[23] LeCun, Y., Bengio Y. e Hinton, G. "Deep learning". En: *Nature* 521 (2015), 436–444.

[24] Jianxin Wu. "Convolutional neural networks". En: *National Key Lab for Novel Software Technology Nanjing University, China* (2020).

[25] Ross Girshick y col. "Region-based Convolutional Networks for Accurate Object Detection and Segmentation". En: *UC Berkeley* ().

zantes de diferentes tamaños para ubicar objetos, las cuales son integradas recursivamente para definir regiones propuestas. Cada región propuesta es mapeada a una arquitectura convolucional para obtener un embebido compacto de representación. Finalmente, se entrena una máquina de soporte vectorial (Support Vector Machine (SVM) [26]) con los embebidos resultantes para clasificar cada una de las regiones, así como también se integra un modelo de regresión para definir las fronteras en la región de interés. En la Figura 2 se ilustra un esquema de esta arquitectura. A pesar que esta propuesta demostró resultados sobresalientes con respecto al estado del arte en imágenes naturales, los mecanismos de propuesta de regiones son computacionalmente costosos. Además, la arquitectura CNN involucra tiempos adicionales de entrenamiento, y su tarea de discriminación está sujeta a realizar predicciones únicamente por los candidatos entregados por una metodología externa. En cuanto al problema particular abordado en este proyecto, el ajuste del mecanismo para selección de regiones candidatas puede ser exhaustivo, además el costo computacional puede elevarse significativamente teniendo en cuenta las dimensiones convencionales de una placa histológica. Por otra parte, en este enfoque no se incluye el modelamiento de regiones no paramétricas, al realizar la representación de detección únicamente en forma de ventanas rectangulares o cuadros delimitadores, lo cual puede ser limitante para el etiquetado de grados de Gleason.

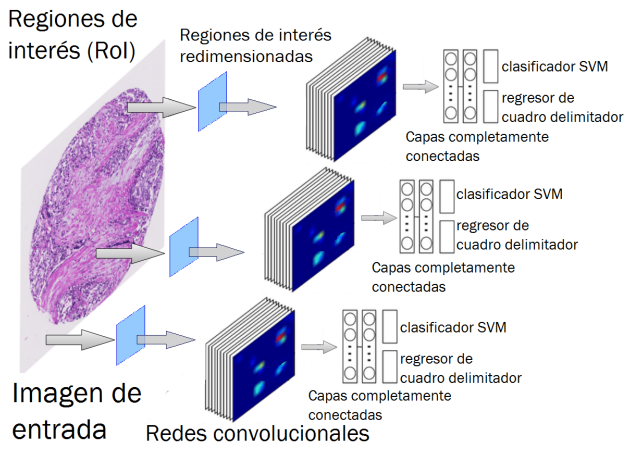
Fast R-CNN: Esta estrategia evoluciona su precedente R-CNN, enfocándose en realizar la operación de convolución solamente una vez por imagen, siendo en consecuencia más rápida computacionalmente [27]. Este método también utiliza la búsqueda selectiva para proponer regiones de interés (RoIs). En la única CNN en donde se procesa toda la imagen, se extrae la parte correspondiente del mapa de características para cada RoI. Cada mapa de características de RoI es redimensionado a un tamaño fijo a partir de una capa de agrupación, y posteriormente

---

[26] Simon Tong y Daphne Koller. "Support Vector Machine Active Learning with Applications to Text Classification". En: *Journal of Machine Learning Research* (2001).

[27] Ross Girshick. "Fast R-CNN". En: *arXiv:1504.08083v2* (2015).

Figura 2. Arquitectura R-CNN. En este método de detección de objetos, el algoritmo de búsqueda selectiva genera múltiples regiones de interés sobre la imagen de entrada. Luego, se implementa una CNN en la parte superior de cada región detectada. La extrapolación de la salida de cada CNN se ingresa en dos vectores de salida: una SVM para clasificar la región y un regresor de cuadro delimitador para corregir la ventana de detección pronosticada.

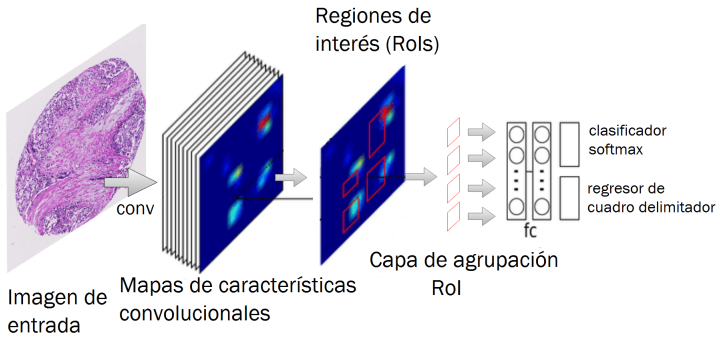


son mapeados en subredes para una clasificación de los objetos por softmax y para obtener una delimitación de la región. En la Figura 3 se representa un esquema de esta arquitectura. Este método es relativamente más eficiente que R-CNN, sin embargo, en términos del reconocimiento de grados histológicos, este enfoque puede tener limitaciones al proponer regiones únicas limitadas a las regiones salientes de la búsqueda selectiva. En este caso, se podrían localizar patrones en donde más se centraliza un grado de Gleason particular, pero perdiendo de vista regiones de un mismo grado y alejadas del patrón de atención.

Faster R-CNN: Esta estrategia que se fundamenta en las dos arquitecturas previas, logra reducir el costo computacional de forma significativa, reemplazando el mecanismo de selección de regiones propuestas. Para ello, la Faster R-CNN utiliza un enfoque *"end-to-end"*, reemplazando el componente basado en grafos para generar regiones con potenciales objetos, por una representación convolucional [28]. En este caso se propone una red denominada RPN (por sus

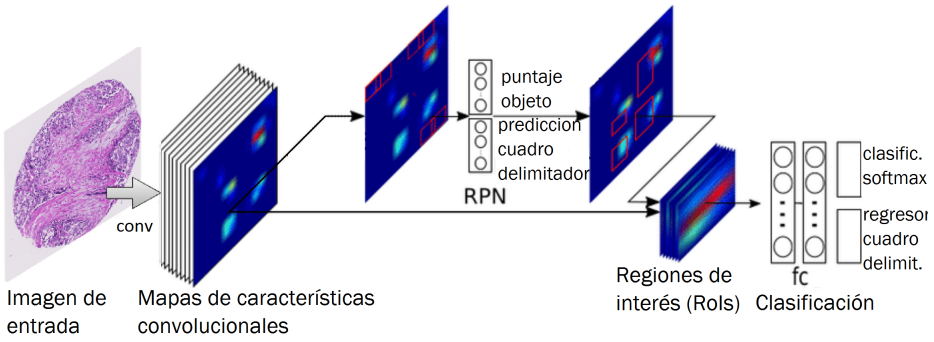
[28] Shaoqing Ren y col. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". En: *arXiv:1506.01497v3* (2016).

Figura 3. Arquitectura Fast R-CNN. En este método, una imagen de entrada y múltiples regiones de interés (RoI) se ingresan en una red totalmente convolucional. Cada RoI es agrupada en un mapa de características de tamaño fijo y luego se asigna a un vector de características por capas completamente conectadas (FC). La red tiene dos vectores de salida para cada RoI: probabilidades de clasificación dadas por softmax y compensaciones de regresión de cuadro delimitador por clase.



siglas en inglés: *regional proposal network*). Esta red propone regiones sobre un conjunto de activaciones de alto nivel, siendo las regiones más salientes las asociadas con objetos. De esta manera, se recuperan un conjunto de activaciones en la última capa, las cuales son tratadas como mapas con focos específicos de atención relacionados con los objetos. Esta propuesta utiliza un mecanismo de ventana deslizante para la detección y clasificación ágil de objetos. Este proceso, además de ser eficiente, logra mapear densamente el mapa de activaciones para hacer una recuperación densa de los objetos de interés. Las regiones de interés (RoIs), extraídas de los mapas resultantes en la RPN, son luego utilizadas para resolver un problema de clasificación (etiqueta de la región) y un problema de regresión (índices espaciales de la región). La Figura 4 muestra un esquema de la arquitectura de este método. Una principal limitación de esta arquitectura para el problema de interés de este trabajo, es su entrenamiento basado en regiones rígidas y parches regionales fijos que demarcan los objetos de interés. Esta representación no resulta natural en el problema de estratificación de Gleason, teniendo en cuenta las regiones que demarcan un estadío de Gleason particular. Además, la conjunción de diferentes arquitecturas para proponer y definir los objetos pueden sumar errores en la estimación.

Figura 4. Arquitectura Faster R-CNN. Faster R-CNN consiste en que sobre el mapa de características de una CNN inicial, RPN actúa para generar múltiples regiones. Cada región consiste en un puntaje y las coordenadas de la ventana. En cada RoI seleccionada se resuelve un problema de clasificación y de regresión para el cuadro delimitador.



R-FCN (Redes totalmente convolucionales basadas en regiones): Para el desarrollo de esta arquitectura, se observó de las propuestas previas, que la capa densa (*fully connected (FC)*) que opera sobre las regiones propuestas para el proceso de clasificación e identificación de fronteras de la región de interés, tiene un costo computacional elevado. La naturaleza de esta capa además incrementa su costo computacional polinómicamente a medida que se adicionan nuevas neuronas, debido a su correlación con entre todas las activaciones [29].

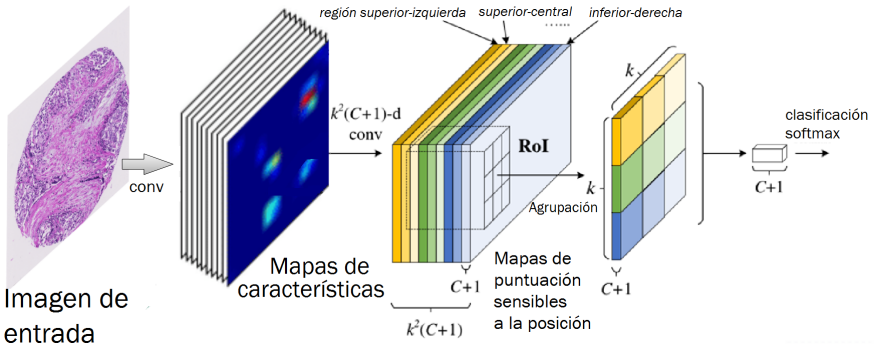
Entonces, la arquitectura R-FCN propone aprovechar la jerarquía matricial de la red RPN y sobre esta hacer cálculos de predicción sobre cada uno de los píxeles de la activación. Para ello se debe preservar la proyección espacial de los píxeles, teniendo cada uno de ellos una referencia de las C clases por reconocer en la escena. Entonces, en la última capa se realiza una convolución $(k \times k)$(C+1)-d, en donde para cada clase C+1 (C es el número de clases más la clase de fondo), habrá $(k \times k)$ mapas de características. Este mapa de características o mapa de puntaje sensible a la posición $(k \times k)$, es utilizado para luego usar un promedio de votación y hacer la anotación sobre una respectiva RoI. De esta manera, todas las regiones propuestas después de la agrupación RoI, utilizan un mismo conjunto de mapas de puntuación (que incluye

[29] Jifeng Dai y col. "R-FCN: Object Detection via Region-based Fully Convolutional Networks". En: *arXiv:1605.06409v2* (2016).

las probabilidades de que la RoI contenga el objeto para cada subregión). También se realiza la regresión de cuadro delimitador para cada objeto. La Figura 5 representa la arquitectura de R-FCN.

A diferencia de los métodos anteriores (R-CNN, Fast R-CNN, Faster R-CNN) que utilizan costosas subredes para cada RoI, esta propuesta utiliza un detector de regiones totalmente convolucional, lo que aumenta la velocidad de entrenamiento y es flexible para adoptar diferentes esquemas convolucionales del estado del arte. Sin embargo, este método depende de los mapas de puntuación sensibles a la posición para determinar la probabilidad de clase del objeto, y realiza un modelamiento de aprendizaje a partir de subregiones para cada región de interés, lo cual puede resultar limitante para determinar regiones que representen estadíos de Gleason, debido a la complejidad en las imágenes histológicas.

Figura 5. Arquitectura R-FCN para la detección de objetos. En esta figura, hay k × k = 3 × 3 mapas de puntaje sensibles a la posición generados por una red totalmente convolucional. Para cada uno de los cuadros k × k en un RoI, la agrupación solo se realiza en uno de los k^2 mapas (marcados por diferentes colores).



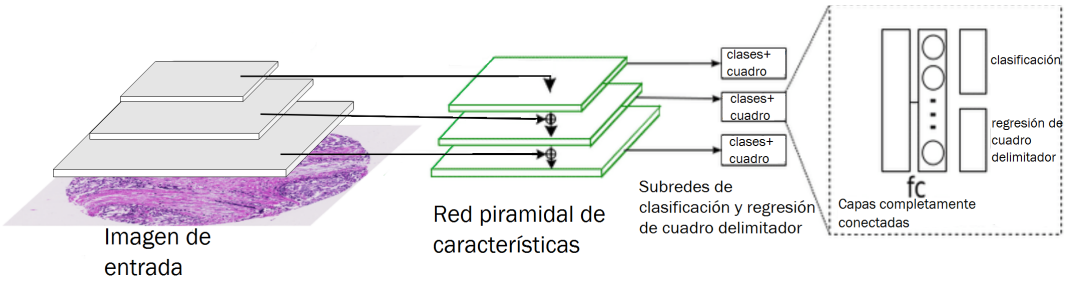
RetinaNet: El método de detección de objetos RetinaNet [30] implementa una Red Piramidal de Características (FPN) para caracterizar objetos a diferentes escalas, permitiendo propiedades de invarianza en la representación de los objetos. La representación piramidal, establecida en cada capa de la arquitectura, se fusiona para, en conjunto, predecir la probabilidad de clasifica-

[30] Tsung-Yi Lin y col. "Focal Loss for Dense Object Detection". En: *arXiv:1708.02002v2* (2018).

ción de objeto en cada localización espacial. También se integra una red de regresión, que hace retropropagar el desplazamiento de los cuadros delimitadores, para cada objeto perteneciente a la verdad fundamental (ver Figura 6).

Además, utiliza una función de pérdida denominada pérdida focal, siendo una modificación de la clásica entropía cruzada que permite enfocarse en ejemplos negativos complejos de representación. Teniendo en cuenta los escenarios típicos de reconocimiento y localización de objetos, esta modificación en la función de perdida permite hacer un balance entre las clases negativas y positivas del conjunto de datos. RetinaNet es un método de una sola etapa, que ha demostrado buenos resultados y una óptima precisión para la detección de objetos en imágenes naturales, aéreas y satelitales, que poseen objetos densos y en pequeñas escalas. Para el problema presentado en el presente trabajo, existe una limitación en cuanto a la delimitación y representación de falsos positivos. En el caso de histología, el desbalance de clases no tiene la misma relación que en las imágenes naturales y su representación multiescala puede perder detalles en las estructuras celulares, que resultan fundamentales para definir grados de Gleason. Además, la salida de este método son cuadros de anclaje para cada región detectada, lo que resulta insuficiente al ser necesaria una representación de segmentación regional que determine los grados de Gleason a nivel de píxel.

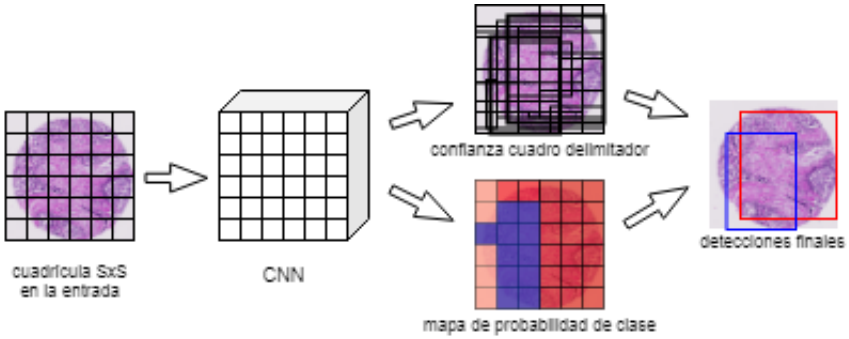
Figura 6. Arquitectura RetinaNet. La arquitectura de red RetinaNet utiliza una FPN sobre una arquitectura ResNet para poder generar una pirámide convolucional multiescala. A esta columna vertebral, RetinaNet adjunta dos subredes. Una subred es para clasificar los cuadros delimitadores y la otra para hacer una regresión de cuadros delimitadores.



YOLO: Esta arquitectura ha marcado una tendencia en cuanto a las representaciones profundas para el reconocimiento y localización de objetos. A diferencia de las arquitecturas previas, esta red utiliza una única red convolucional para predecir los cuadros delimitadores y las probabilidades de clase para estos cuadros [31]. Esencialmente, la referencia espacial de una región en la imagen es proyectada a través de las convoluciones, permaneciendo su referencia hasta la última capa de predicción. Para ello, la imagen de entrada se divide en una cuadrícula $S \times S$, y tomando un conjunto de m cuadros delimitadores. Para cada cuadro delimitador entonces se genera la probabilidad de clase y los valores de compensación de cada región. En la Figura 7 se puede observar un esquema de esta arquitectura.

Sin embargo, YOLO tiene una limitación debido al número de cuadros delimitadores, y está sesgada por una búsqueda en cuadrícula que es trazada al principio de la imagen. En lo que respecta al problema presentado en el presente trabajo, las imágenes histológicas son imágenes que no se pueden limitar a cuadros delimitadores y cuadrículas. A pesar de los resultados probados en la literatura, esta arquitectura restringe la delimitación a una grilla inicial de configuración, lo cual no resulta natural en el problema de segmentación de patrones de Gleason.

Figura 7. Arquitectura YOLO. El sistema realiza y modela la detección como un problema de regresión. Esto se hace dividiendo la imagen en una cuadrícula de tamaño S × S. Posteriormente, para cada cuadrícula se realiza una predicción de m cuadros delimitadores, la confianza de esos cuadros y las probabilidades de clase.



[31] J. Redmond y col. "You Only Look Once: Unified, Real-TimeObject Detection". En: *arXiv:1506.02640v5* (2016).

2. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

El cáncer de próstata es el cáncer más frecuente en los hombres y la segunda causa principal de muerte por cáncer [32]. El sistema de puntuación de Gleason es el sistema estándar para cuantificar la agresividad de la enfermedad. Este sistema es generado rutinariamente en laboratorios clínicos con buena reproducibilidad. Sin embargo, este proceso conlleva mucho tiempo (aproximadamente uno a tres días para etiquetar 6 a 15 muestras de una biopsia), y es totalmente dependiente de la interpretación y capacidad observacional del experto. Es por ello que se presenta una persistente variabilidad en el diagnóstico de la enfermedad, evidenciando la concordancia entre observadores (según el valor kappa ponderado) desde valores máximos de 0.55, hasta valores mínimos de 0.33, especialmente para grados intermedios de la escala de Gleason (3 y 4). Estudios recientes se centran en el análisis y la caracterización local de la enfermedad basada en parches, lo que en muchos casos no tiene una apropiada reproducibilidad clínica al no estratificar de forma global la muestra histológica, y al no tener en cuenta en ella zonas de interés y patrones espaciales de forma completa que poseen relevancia en el diagnóstico final.

[32] American Society of Clinical Oncology. *Cáncer de próstata: Estadísticas*. En: Cancer.Net. 2018.

3. OBJETIVOS

Objetivo general

Desarrollar una estrategia convolucional profunda para la segmentación y estratificación de patrones relacionados con la severidad del cáncer de próstata.

Objetivos específicos

- Seleccionar un conjunto de datos etiquetado público de imágenes histológicas del cáncer de próstata en la escala de Gleason.
- Identificar esquemas de segmentación de objetos supervisada en imágenes histológicas.
- Desarrollar una estrategia de segmentación de objetos supervisada basada en aprendizaje profundo que permita una diferenciación entre grados de severidad cáncer de próstata.
- Evaluar la estrategia de segmentación propuesta utilizando métricas relacionadas con la caracterización de grados de cáncer según la escala de Gleason.

4. REPRESENTACIÓN PROFUNDA MULTINIVEL PARA LA SEGMENTACIÓN REGIONAL SEGÚN ESCALA DE GLEASON

Los patrones espaciales que demarcan un grado específico de Gleason agrupan estructuras histológicas, que pueden ser regiones características de un estadío particular de la enfermedad. Bajo esta hipótesis, en este trabajo se propuso un método de segmentación para modelar las anotaciones de grados de Gleason, pretendiendo ser una herramienta de soporte para la rutina de análisis de los patólogos. El esquema de segmentación propuesto utiliza como módulo de modelamiento una representación profunda de tipo Mask R-CNN, ya que ha demostrado robustez para obtener segmentaciones semánticas en diferentes escenarios. Teniendo en cuenta la complejidad del problema de caracterización de grados de severidad de cáncer, el carácter global de Mask R-CNN y la carencia en la definición de regiones locales en la tarea histopatológica, aquí se implementó una estrategia progresiva que entrena diferentes versiones de Mask R-CNN para obtener una segmentación final. En el primer nivel de representación, se entrenó una red regional utilizando anotaciones delineadas por un experto patólogo en imágenes histológicas. En el segundo nivel de representación, fue entrenada una red con anotaciones que reportan una probabilidad alta, dada por el primer nivel de representación, para diversos grados de Gleason. Una última estimación de los grados de Gleason fue recuperada desde las activaciones convolucionales de una representación profunda, que indican patrones de atención dados por la Mask R-CNN para la localización de regiones de interés en placas histológicas, extrayendo nuevos indicadores y definiendo nuevas regiones de forma local. El consenso de las tres segmentaciones da origen a la segmentación propuesta para soportar la tarea de análisis de placas histológicas. La metodología propuesta se detalla en la Figura 8.

Figura 8. Estrategia propuesta. Se utilizó una estrategia multinivel de aprendizaje profundo, utilizando tres representaciones de tipo Mask R-CNN. El primer nivel toma muestras histológicas completas y genera una primera máscara de segmentación global. El segundo nivel fue alimentado por las regiones que corresponden a la intersección de las máscaras predichas sobrelapadas, generadas por el primer nivel. El tercer nivel de representación regional fue entrenado a partir de las regiones resultantes de las activaciones convolucionales de la primera red. Estas activaciones fueron sumadas transversalmente mediante SumPooling y binarizadas para obtener máscaras de entrenamiento para el tercer nivel. El resultado final se dió por la superposición de la salida de los tres niveles de representación.

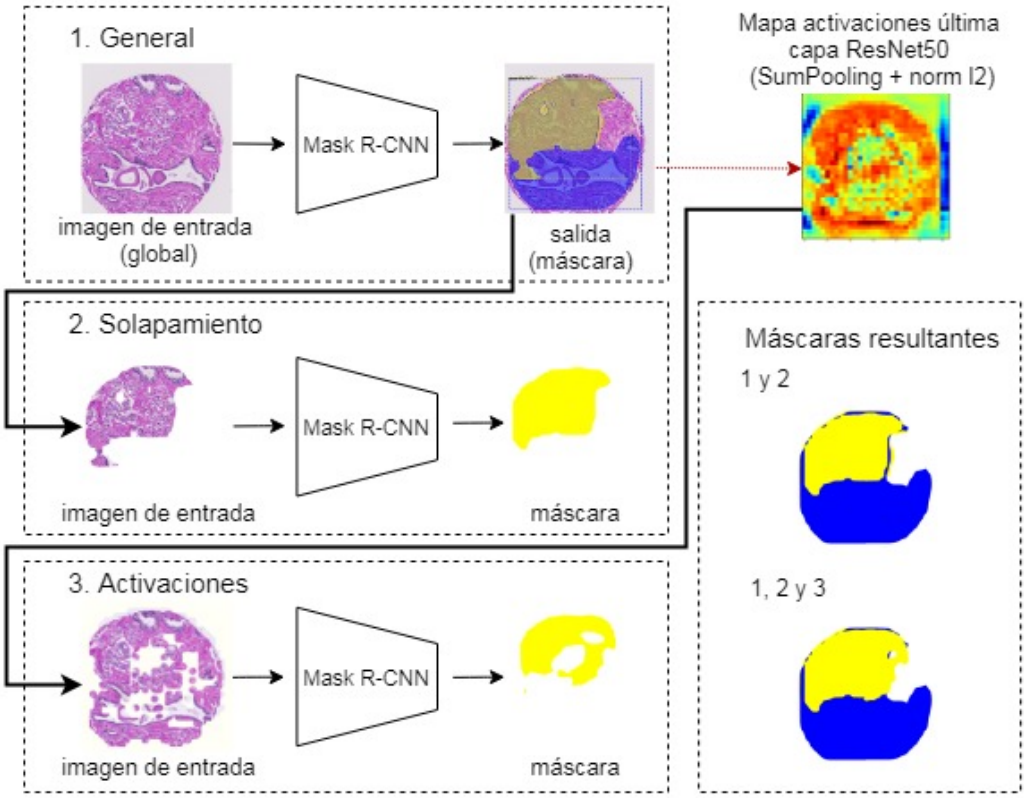
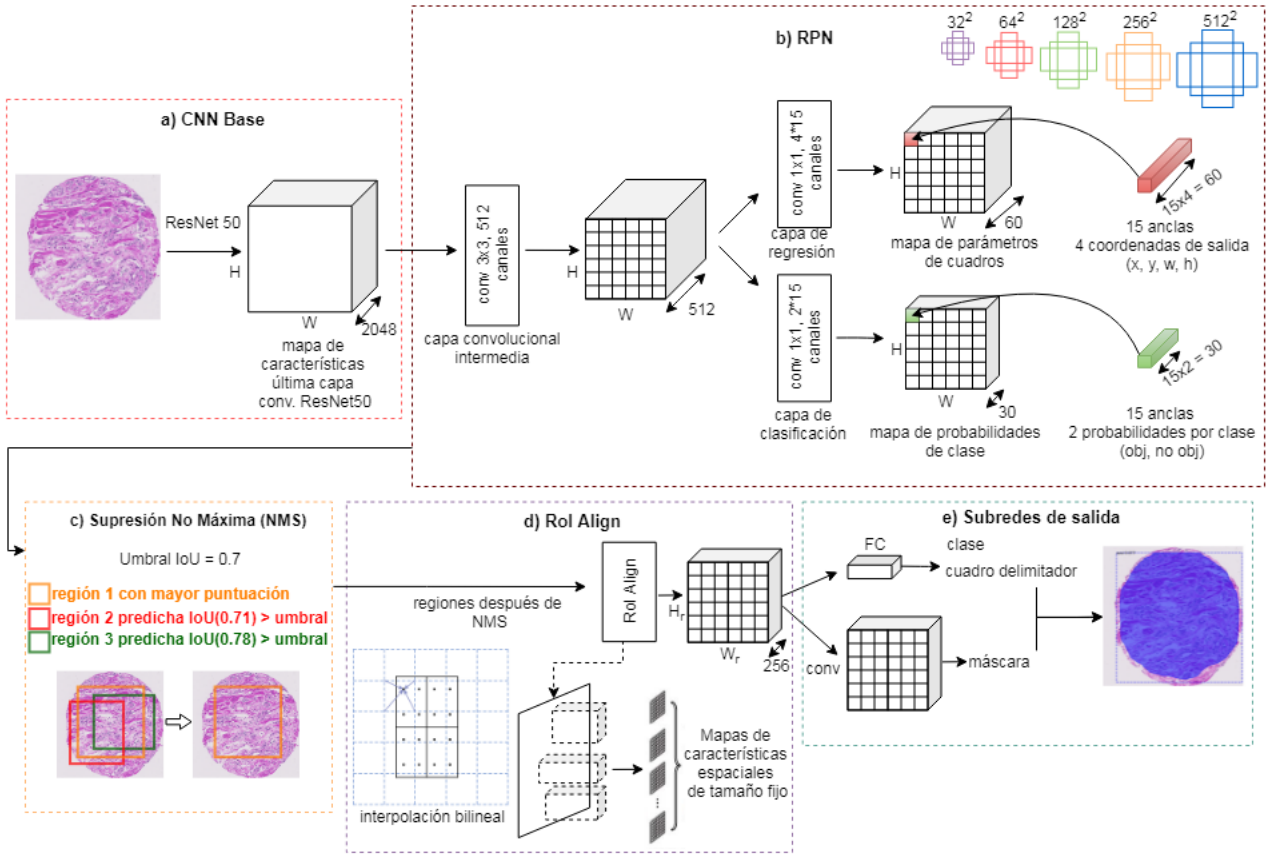


Figura 9. Método Mask R-CNN para la segmentación de instancias.



4.1. Unidad regional de representación: Mask R-CNN

Con el fin de obtener una segmentación regional, en este trabajo se utiliza como unidad de representación la arquitectura Mask R-CNN, entrenada con diferentes conjuntos de datos y utilizando activaciones intermedias como pseudo-segmentaciones, que guían la delineación final propuesta en la metodología. La red Mask R-CNN es una variante de los modelos R-CNN, previamente descritos en el marco teórico, como la *Fast R-CNN* y la *Faster R-CNN*. También, la Mask R-CNN es una representación profunda que constituye el estado del arte en algunos dominios de visión por computador, permitiendo obtener máscaras resultantes que delimitan

objetos de aprendizaje, es decir, realiza una segmentación semántica [33]. En la Figura 9 se ilustra el esquema del funcionamiento de esta arquitectura. Para realizar una explicación a detalle, se hizo una división en 5 etapas: entrada a la CNN base, RPN, Supresión No Máxima, RoIAlign y subredes de salida, las cuales serán descritas a continuación.

CNN Base. En la primera fase, las imágenes de entrada se mapean a una representación profunda. Particularmente en este trabajo, la placa histológica de entrada tiene dimensiones relativamente grandes ($3100 \times 3100 \times 3$), la cual codifica patrones en múltiples escalas espaciales y relaciones biológicas complejas que deben representarse densamente a través de arquitecturas suficientemente profundas. Sin embargo, una principal limitación para entrenar redes profundas, en el escenario histológico, es el limitado número de muestras para lidiar con la amplia variabilidad y las relaciones visuales complejas que se representan. Es por ello que en este trabajo, como base de representación visual convolucional, se decidió utilizar una arquitectura residual ResNet-50. Esta arquitectura implementa conexiones residuales entre capas de la arquitectura, que permite afrontar problemas de desvanecimiento del gradiente durante el entrenamiento y logra un entrenamiento relativamente estable con un conjunto de datos compacto. Esta red logra aprender jerárquicamente patrones visuales de los estadíos de Gleason y resaltar las principales características que pueden estar correlacionadas con la severidad de la enfermedad. Las imágenes de entrada fueron redimensionadas a $1024 \times 1024 \times 3$ para el entrenamiento, una dimensión que sigue siendo relativamente grande, pero suficiente para acelerar el tiempo de ejecución al momento de entrenar la representación profunda. Esta fase es ilustrada en la Figura 9-a.

RPN (*Regional proposal network*). Una vez entrenada la representación convolucional profunda, las capas codifican características de diferente orden de representación, que pueden tener una relación asociada con las anotaciones de entrenamiento. Entonces, una representación

---

[33] Kaiming He y col. "Mask R-CNN". En: *arXiv:1703.06870v3* (2018).

de alto nivel (capa cercana a la predicción) es utilizada para realizar la cuantificación de regiones de interés, y su posterior segmentación. Específicamente, las activaciones se extraen de la última capa (en este caso, de dimensiones $32 \times 32 \times 2048$). Posteriormente, estas activaciones son nuevamente convolucionadas, en el marco Mask R-CNN, por una red donde se extraen patrones visuales que se denomina red de propuesta de regiones (RPN, *region proposal network* por sus siglas en inglés). La red entrenada sobre esta capa tiene como único objetivo generar posibles regiones que contengan un objeto, llamadas regiones propuestas. El mapa de características de la última capa es la entrada a la red RPN. Sobre este mapa, se implementa una capa convolucional intermedia con una ventana deslizante de tamaño 3×3 y con 512 canales o filtros, cuyo propósito es extraer mapas de características para realizar una identificación y ubicación de las regiones, y poder generar regiones candidatas. El resultado de ello es un mapa de características con las mismas dimensiones de anchura y altura, y con una profundidad de 512 canales, que es posteriormente ingresado en dos capas convolucionales adicionales, con convoluciones de 1×1 y con paso de 1, encargadas de realizar una regresión y una clasificación de las regiones.

Específicamente para cada posición de la ventana deslizante, la RPN puede predecir múltiples propuestas de región delimitadas por rectángulos de anclaje. Estos rectángulos de anclaje están centrados en cada posición de la ventana deslizante de las capas anteriormente mencionadas, y consisten en diversas escalas que indican tamaños de la región, y diferentes relaciones de aspecto que indican proporciones entre la anchura y la altura de una región. Una ilustración de los rectángulos de anclaje están disponibles en la Figura 9-b. Por ejemplo, en este trabajo se utilizó RPN para proponer RoIs en 5 escalas o tamaños diferentes (32×32, 64×64, 128×128, 256×256 y 512×512) y 3 relaciones de aspecto (1:2, 1:1 y 2:1) para cada uno de los tamaños. De esta forma, RPN predice 15 regiones candidatas en cada posición de la ventana deslizante que corresponden a esas 15 anclas, y se generan $H \times W \times no.anclas$ regiones (específicamente $32 \times 32 \times 15$) que pueden contener un objeto en específico. Entonces, en la última capa existe un modelo de regresión encargado de predecir las coordenadas de las propuestas, y tiene 4×15 canales, siendo 15 el número anclas y 4 el número de coordenadas a devolver para cada región

propuesta: *x, y, w* y *h* , en donde x y y representan el centro de la región y w y h representan la anchura y la altura, respectivamente. También, de la última convolución se sobrepone una capa de clasificación, encargada de devolver una probabilidad de contener un objeto en específico. Posee 2×15 canales de profundidad, siendo 15 igualmente el número de anclas y 2 la probabilidad binaria de que contenga y no contenga un objeto en la región. De esta manera, cada región propuesta por RPN devuelve una puntuación relativa a la presencia del objeto en esa región concreta y las coordenadas que representan la región propuesta. El entrenamiento de RPN se realiza mediante retropropagación y teniendo en cuenta la intersección sobre la unión (IoU), entre el cuadro delimitador predicho y un cuadro delimitador de *ground-truth*, etiquetando una RoI como positiva si IoU es mayor a 0.7, y negativa si es menor. De esta manera se asigna mayor probabilidad de contener un objeto a aquellas regiones positivas.

Supresión No Máxima (*NMS*). La RPN genera miles de regiones, debido a las múltiples regiones establecidas a partir de las compensaciones para cada ancla, en cada posición de la ventana deslizante en el mapa de características, de tal manera que RPN predice múltiples cuadros delimitadores para una misma instancia. Sin embargo, procesar esto en la red es un trabajo complejo, debido al gran número de propuestas de regiones (15.000 aproximadamente). Por tanto, para evitar reiteradas detecciones para una misma instancia, para todas las regiones predichas por RPN se realiza un cálculo para seleccionar solamente un conjunto de regiones significativas. Este cálculo es basado en la intersección sobre la unión (IoU), entre un cuadro delimitador predicho para una instancia que tenga una puntuación de confianza alta (probabilidad de que contenga un objeto), y los demás cuadros delimitadores. Dado un umbral experimental establecido de 0.7 de IoU, las RoI mayores a ese umbral se omiten, considerando solamente el mejor cuadro delimitador de esa instancia para la siguiente etapa e ignorando los restantes, debido a que poseen un sobrelapamiento con el cuadro seleccionado. Esta técnica se denomina Supresión No Máxima (NMS). Una ilustración de esta fase está disponible en la Figura 9-c.

Alineamiento de las regiones propuestas (*RoIAlign*). La RPN devuelve propuestas de región presentando sus coordenadas en función de la imagen original. Posteriormente, cada coordenada es normalizada para representarla en función del mapa de características utilizado (de la última capa convolucional de la CNN Base). Así, se obtienen nuevas coordenadas de las regiones, relativas al tamaño del mapa de características, para poder generar los mapas de características espaciales. Un mapa de características espacial es la parte correspondiente, en el mapa de características, de cada región que contiene un objeto (RoI). Teniendo en cuenta que las dimensiones propuestas (en el espacio de la imagen) y su proyección en las activaciones (convolucionadas a través de diferentes capas) difieren en tamaño, es necesario alinear estas regiones para encontrar los descriptores correspondientes de cada RoI. Por lo tanto, se utiliza una operación para cada RoI llamada RoIAlign (alineación de región de interés). RoIAlign consiste en dividir el mapa de características en una cuadrícula de k celdas, en donde para definir valores concretos, usando interpolación bilineal, cada celda es subdividida en 4 puntos que contienen el valor correspondiente del píxel más cercano en el mapa de características. De estos 4 valores se obtiene el valor promedio correspondiente a cada celda de la cuadrícula. Esta operación devuelve todas las RoI en un tamaño de imagen fijo para la red de detección de objetos, y aproxima la desalineación para que las RoI puedan asignarse con mayor precisión a las regiones de la imagen original. Los resultados de la alineación de las regiones de interés fueron mapas de características espaciales de tamaño $7 \times 7 \times 256$. Una ilustración de esta fase es ilustrada en la Figura 9-d.

Subredes de salida. Finalmente, para cada RoI alineada, se resuelve un problema de clasificación (etiqueta de la región), un problema de regresión de caja delimitadora (índices espaciales de la región) y una rama para la predicción de máscara de segmentación. Las subredes de clasificación de objeto y regresión de cuadro delimitador están compuestas por capas completamente conectadas (FC) y la subred de predicción de máscara es una pequeña red completamente convolucional, compuesta por cuatro capas de convolución consecutivas de 3×3, una capa de deconvolución 2×2 y una capa final de convolución de 1×1. La subred de predicción de máscara

genera máscaras de segmentación de dimensión $m \times m$ para cada clase y cada RoI. Teniendo K máscaras en total, la salida de esta subred tiene un tamaño de $K \times m^2$. Las máscaras de segmentación generadas son máscaras binarias de tamaño 1024×1024, con valores de 1 en las ubicaciones donde el píxel contiene el objeto, y valores de cero donde no lo contiene. Esta fase es ilustrada en la Figura 9-e.

La minimización y aprendizaje *end-to-end* de la arquitectura Mask R-CNN se realiza mediante la convergencia de diferentes funciones de pérdida. Para las tareas de clasificación, regresión y predicción de máscara, la Mask R-CNN emplea la siguiente pérdida de tareas múltiples:

$$L(p_i, t_i) = \underbrace{\frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*)}_{L_{cls}} + \underbrace{\lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)}_{L_{reg}} + \underbrace{\frac{1}{N_{mask}} \sum_i L_{mask}(p_i^{pixel}, p_i^{pixel,*})}_{L_{mask}} \quad (1)$$

En donde p_i es la probabilidad de clasificación para cada predicción i, p_i^* es un valor de 1 para anclas positivas y 0 para anclas negativas, N_{cls} es un número de anclas en un mini-lote escogidas para la tarea de clasificación, N_{reg} es el número total de anclas, λ es un valor constante, t_i es la caja predicha y t_i^* es la caja de *ground-truth*, representadas en términos de coordenadas de cuadro delimitador. L_{cls} es una función de pérdida logarítmica $L_{cls}(p, u) = -logp_u$ y L_{reg} es definida como una función de pérdida *smooth-L1*, $L_{reg} = smooth_{L1}(t_i - t_i^*)$. La Mask R-CNN también añade la pérdida de la tarea de predicción de máscara como una pérdida L_{mask} de entropía cruzada binaria promedio, entre las máscaras de *ground-truth* y las máscaras predichas. L_{mask} solo es definido en un número de máscaras N_{mask} en donde la región segmentada está asociada con la clase de *ground-truth*. Esta pérdida L_{mask} opera de forma densa en cada uno de los píxeles envueltos dentro de una región positiva.

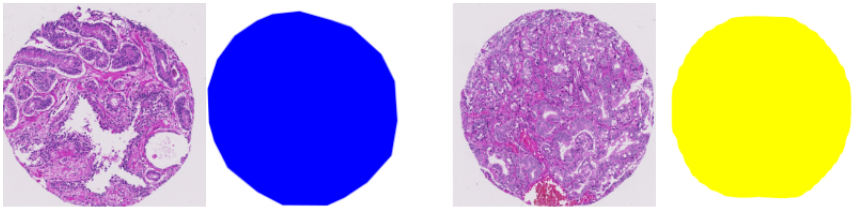
4.2. Primer nivel: Representaciones primarias basadas en anotaciones.

En el primer nivel de representación, para todo el conjunto de datos de entrenamiento, cada imagen de entrada (imagen histológica completa) se introduce en un primer modelo Mask R-CNN que se denominó Mask R-CNN General. Esta red se entrenó utilizando las anotaciones completas realizadas por un único experto patólogo, capturando de esta manera patrones globales de la muestra histológica, por lo que su resultado fueron máscaras globales con anotaciones de áreas asociadas a un puntaje de Gleason. Esta unidad Mask R-CNN se entrena en un esquema multiclase (4 clases), lo que permite generar segmentaciones según los grados de gleason más probables dados los patrones visuales y las segmentaciones dadas en el entrenamiento. En términos generales, este primer nivel de segmentación logra obtener segmentaciones globales, con una buena asociación, cuando solo existe un grado de Gleason en toda la placa histológica (la Figura 10 muestra un ejemplo de predicción de máscara). Por lo anterior se puede inferir que esta representación no es suficiente para diferenciar regiones pequeñas y para detectar patrones locales en imágenes histológicas que puedan contener distintos grados de Gleason. Por esta razón es necesario definir estrategias complementarias que permitan redefinir regiones de máscaras globales, o que puedan ubicar otras asociaciones de grados diferentes en segmentos más pequeños de la imagen. Por ejemplo, una de las características de este nivel es la generación de múltiples máscaras sobrelapadas y con un alto nivel de probabilidad. Estas asociaciones sobrelapadas fueron determinadas como las más desafiantes, por lo cual, se propuso un nuevo nivel de segmentación, que involucrara únicamente estas regiones para dar mayor complejidad al conjunto de entrenamiento y forzar a la representación a separar estas segmentaciones. En la siguiente subsección se define la solución propuesta para estos casos de solapamiento de etiquetas.

4.3. Segundo nivel: Representaciones específicas en regiones de frontera de Gleason.

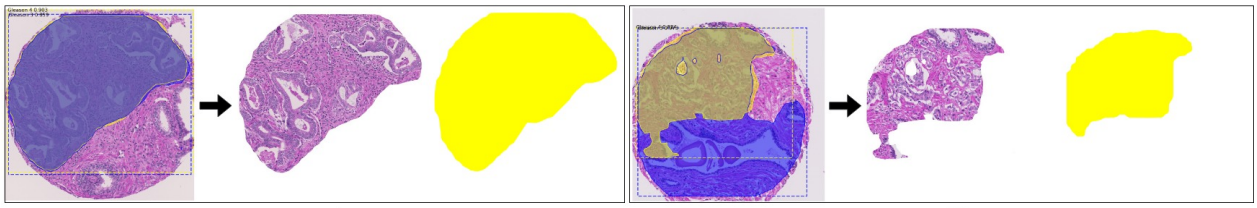
En algunos casos, la salida generada por el primer nivel de representación son dos máscaras superpuestas, es decir, dos máscaras que etiquetan una región específica con dos grados diferentes

Figura 10. Ejemplos para el primer nivel de representación. a) Ejemplo de dato de entrenamiento: imagen histológica completa con su respectiva máscara de anotación (azul - Gleason 3) realizada por un experto patólogo. b) Ejemplo de dato de prueba: imagen histológica completa y su máscara predicha (amarillo - Gleason 4).



de Gleason. Por lo tanto, se implementó un segundo modelo para dar solución a estos casos de solapamiento y para decidir la etiqueta correspondiente a esas regiones (ver Figura 11). Para el entrenamiento de esta segunda red, se realizó una extracción de los fragmentos de la imagen histológica y su máscara, relativos a la región de solapamiento entre las dos etiquetas predichas por la primera red. Para la prueba, solo se tomaron las áreas relativas al solapamiento de las imágenes histológicas para predecirlas. Así, este segundo modelo predice y decide la etiqueta correspondiente en aquellas regiones que reportan una alta probabilidad para diferentes estadíos de Gleason.

Figura 11. Ejemplos para el segundo nivel de representación. a) Ejemplo de dato de entrenamiento: máscara resultado del primer nivel de representación con dos etiquetas superpuestas, fragmento de imagen histológica proporcional al solapamiento con su respectiva máscara de anotación (Gleason 3). b) Ejemplo de dato de prueba: máscara resultado del primer nivel de representación con dos etiquetas superpuestas, fragmento de imagen histológica y su máscara predicha por la segunda red.



4.4. Tercer nivel: Redefinición de fronteras según activaciones convolucionales.

Uno de los desafíos predominantes en la segmentación de grados de Gleason tiene que ver con la definición local de las fronteras de la región. A pesar de los dos niveles de de segmentación definidos previamente, se pudo observar carencia local en la definición de algunas regiones, esto debido al carácter global de la Mask R-CNN. Por lo tanto, como un tercer nivel de procesamiento, se exploraron activaciones y representaciones intermedias de la representación profunda para extraer nuevos indicadores locales para modificar las segmentaciones de un grado de Gleason específico.

Las redes neuronales convolucionales detectan y extraen, en cada capa, características específicas de los objetos presentes en las imágenes, extrayendo así características más complejas en capas más profundas. Ciertas características se resaltan en cada capa convolucional, y corresponden al impacto que tienen en los patrones visuales de una imagen, que logran que las neuronas se activen con mayor magnitud. Así, las activaciones son asociadas a las regiones en donde, para cierta capa convolucional, existe mayor significancia en las imágenes.

Específicamente en este trabajo, se llevó a cabo un tercer nivel de representación Mask R-CNN, en donde se tuvieron en cuenta las activaciones de la última capa de la red convolucional base (mostrada en la Figura 9-a). Esta última capa proporciona información sobre los patrones en los que se enfoca el método Mask R-CNN para localizar las regiones y predecirlas, y donde se centra la RPN para proponer regiones de interés. La Figura 12 muestra algunos ejemplos de estas activaciones. Esta capa tiene un total de 2048 activaciones de tamaño 32×32. Las activaciones de la última capa de la ResNet-50 del primer modelo, se agruparon mediante SumPooling, y las características sumadas se normalizaron en l2 [34][35]. De este modo, se obtuvo el Mapa de

---

[34] A. Babenko y V. Lempitsky. "Aggregating local deep features for image retrieval". En: *International Conference on Computer Vision (ICCV)* (2015).

[35] Y. Kalantidis, C. Mellina y S. Osindero. "Crossdimensional weighting for aggregated deep convolutional features". En: *arXiv:1512.04065* (2015).

Características Agrupadas (MCA). A continuación, en el MCA, se tomaron los píxeles superiores a un umbral de 0.7, y se tomaron las regiones proporcionales a ese umbral para las imágenes histológicas y sus máscaras para el entrenamiento (ver Figura 13). Con estas regiones se entrena un nuevo nivel Mask R-CNN permitiendo una descripción más granular de las regiones. Para la prueba, sólo se tomaron las regiones proporcionales al umbral de activaciones en las placas histológicas, prediciendo así una nueva máscara, determinada a partir de las regiones de las activaciones convolucionales de la primera red.

Figura 12. Ejemplos de 8 mapas de activaciones de la última capa ResNet-50 del primer nivel de representación. En total, esta capa posee 2048 activaciones. La agrupación de estos mapas mediante SumPooling seguido de una normalización en l2, da como resultado el mapa de características agrupadas (última columna a la derecha).

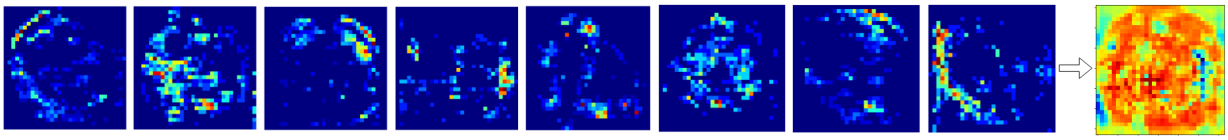
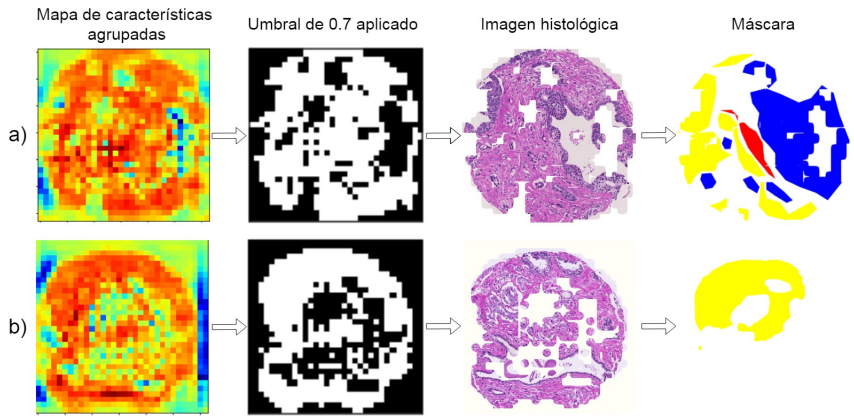


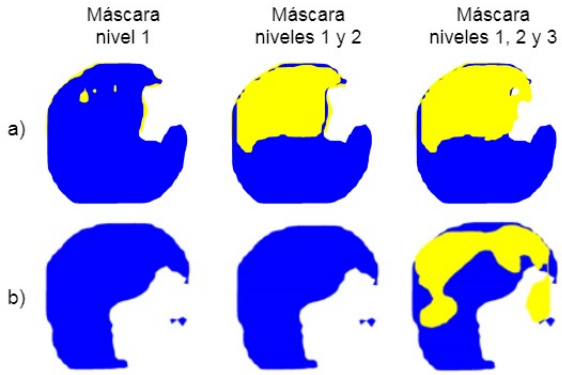
Figura 13. Ejemplos de datos para el tercer nivel de representación. Umbral aplicado en el MCA para un dato de entrenamiento y de prueba. a) Ejemplo de dato de entrenamiento: fragmento de imagen histológica con su respectiva máscara de anotación, proporcionales al umbral aplicado de 0.7 en el MCA obtenido a partir de las activaciones de la última capa ResNet-50 de la primera red. b) Ejemplo de dato de prueba: fragmento de imagen histológica relativa al umbral y su máscara predicha por la tercera red.



4.5. Fusión de representaciones regionales.

El presente trabajo se limitó a realizar una fusión de los resultados de las máscaras de anotación, obtenidas en los diferentes niveles, a nivel lineal. Es decir, se implementó una fusión tardía de las máscaras obtenidas en cada nivel de representación. Sin embargo, se pueden pensar múltiples maneras de explorar representaciones multinivel y fusión de máscaras predichas. La fusión de máscaras generadas en cada nivel de representación se realizó de la siguiente manera: la máscara generada por el segundo nivel de representación se superpone en la salida anterior (máscara global del primer nivel), generando una segunda máscara global a partir de la fusión de los resultados de las dos primeras redes. La máscara de salida predicha por el tercer y último nivel de representación, igualmente se superpone en el resultado anterior, es decir, la máscara acoplada por la predicción de las dos primeras redes, creando una máscara global final, generada a partir de nuevas anotaciones (ver Figuras 14 y 8).

Figura 14. Ejemplo de fusión de representaciones regionales. a) Desde el segundo nivel de representación, para las máscaras predichas por cada nivel se realizaba una superposición de forma lineal, es decir, en la salida de la red anterior. En b), la máscara generada por el nivel 1 y la máscara fusionada por los niveles 1 y 2 es la misma, debido a que no se presentó solapamiento en la primera red para este dato de prueba.



5. DISEÑO EXPERIMENTAL

5.1. Conjunto de datos

La estrategia de segmentación propuesta se entrenó y validó un conjunto de datos, publicado en el repositorio de Hardvard Dataverse [36]. Este conjunto de datos contiene un total de 886 imágenes de muestras de tejido prostático, teñidas con H&E a una resolución de 40x, con un tamaño de 3100 × 3100 píxeles. Cada una de las imágenes fue digitalizada con un escáner de portaobjetos digital Hamamatsu C9600 NanoZoomer 2.0-HT [37]. Cada una de las imágenes tiene su respectiva máscara de anotación según la escala de Gleason, y están agrupadas en 5 microarrays de tejidos (TMAs, por sus siglas en inglés). La asignación de las etiquetas de las regiones fue realizada por un patólogo experto (en este trabajo se denominará patólogo 1) en todo el conjunto de datos y por un patólogo adicional (en este trabajo se denominará patólogo 2) únicamente en el subconjunto de prueba. Cada etiqueta en el dataset tiene un color distintivo para identificar cada clase en la escala de Gleason de la siguiente manera: verde para benigno, azul para Gleason 3, amarillo para Gleason 4 y rojo para regiones con Gleason 5 (ver Figura 15). En cuanto al diseño experimental, los autores del repositorio proponen una partición de los datos, la cual es seguida fielmente en este trabajo. En este caso se utilizaron cuatro TMAs para el entrenamiento con un total de 641 imágenes histológicas y un TMA para la prueba, que corresponde a 245 imágenes histológicas. La distribución de las anotaciones según la escala de Gleason del conjunto de datos se resume en la Tabla 1 y en la Figura 16.

---

[36] Eirini Arvaniti y col. "Replication Data for: Automated Gleason grading of prostate cancer tissue microarrays via deep learning". En: (2018).

[37] Qing Zhong y col. "A curated collection of tissue microarray images and clinical outcome data of prostate cancer patients". En: *Sci Data 4, 170014* (2017).

Figura 15. Ejemplos del conjunto de datos de Hardvard Dataverse. a) y b) representan dos ejemplos de muestras histológicas respecto a cada puntaje, y cada muestra posee encima sus máscaras de anotación etiquetadas por un experto patólogo en la escala de Gleason.

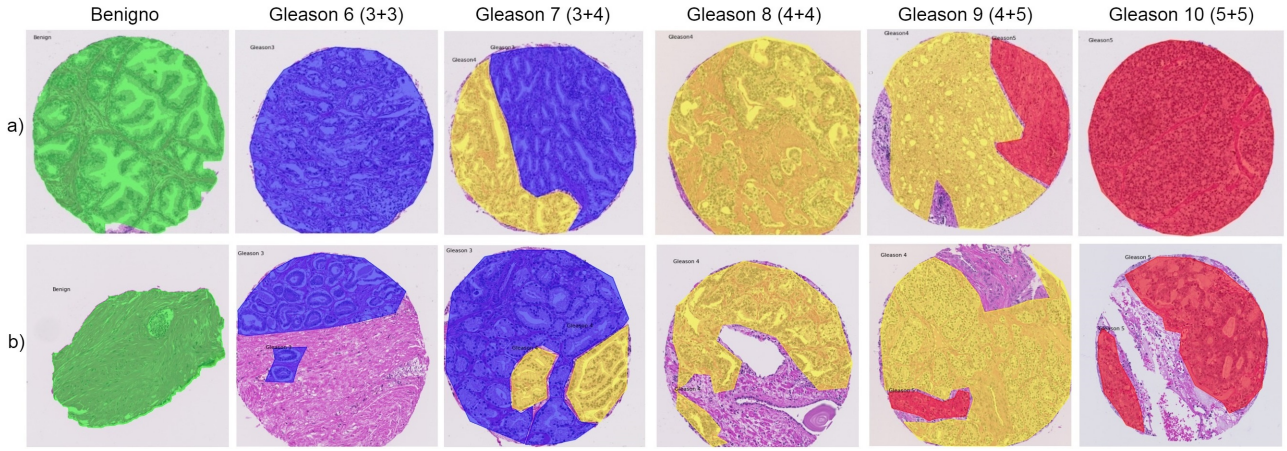


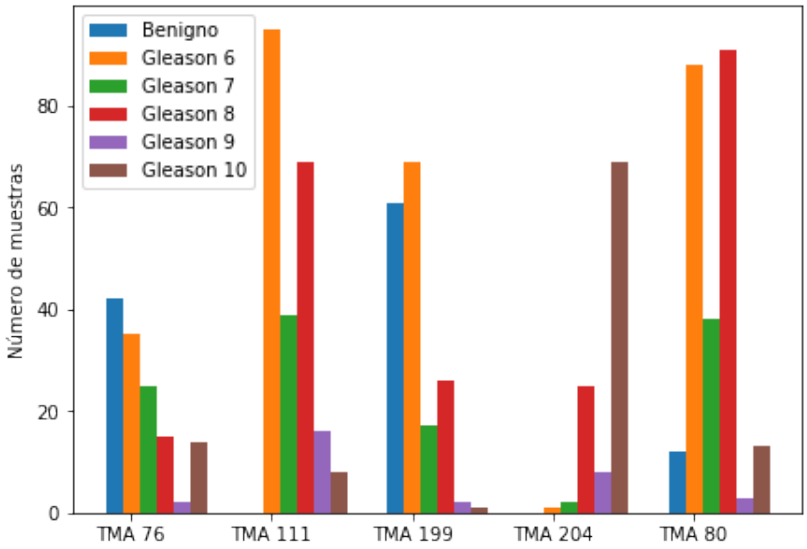
Tabla 1. Distribución de los puntajes de Gleason del conjunto de datos de Hardvard Dataverse. Los puntajes están descritos según la segunda escala de Gleason (suma de los dos grados más predominantes) de la siguiente manera: 6 (3+3), 7 (3+4, 4+3), 8 (4+4, 5+3, 3+5), 9 (4+5, 5+4) y 10 (5+5). TMA 80 se utilizó como conjunto de prueba y los otros TMAs como conjunto de entrenamiento. Todos los TMAs de entrenamiento fueron etiquetados por un patólogo y el TMA 80 por un patólogo adicional

	Benign	6	7	8	9	10	Total de imágenes	Subconjunto
TMA 76	42	35	25	15	2	14	133	
TMA 111	0	95	39	69	16	8	227	Entrenamiento
TMA 199	61	69	17	26	2	1	176	
TMA 204	0	1	2	25	8	69	105	
TMA 80	12	88	38	91	3	13	245	Prueba

5.2. Configuración de la estrategia

Para evaluar el esquema de segmentación propuesto, se compararon primero dos redes como representación visual base de la Mask R-CNN (evaluada en el primer nivel de representación): la ResNet-50 y la ResNet-101. Se utilizó el optimizador determinado por defecto en el marco de la Mask R-CNN, el gradiente descendente estocástico (SGD), variando la tasa de aprendizaje en 0.001, 0.0005 y 0.0001. Para este trabajo se eligió ResNet-50 con una tasa de aprendizaje de 0.001 ya que generó los mejores resultados en resultados preliminares sobre experimentos básicos. La

Figura 16. Gráfico de distribución de los puntajes de Gleason del conjunto de datos. TMA 111 y TMA 199 presentan la mayor cantidad de muestras en el subconjunto de entrenamiento.



RPN se utilizó para proponer RoIs en 5 tamaños (32×32, 64×64, 128×128, 256×256 y 512×512) y en 3 relaciones de aspecto de anclas (1:2, 1:1 y 2:1). El umbral de NMS se fijó en 0.7. En cuanto a las subredes de salida, el regresor de cuadro delimitador y el clasificador de objeto estaban compuestos por capas totalmente conectadas (FC) y el predictor de máscara era una red totalmente convolucional, compuesta por cuatro capas de convolución consecutivas de 3×3, una capa de deconvolución de 2×2 y una última capa de convolución de 1×1 en secuencia. En el proceso de entrenamiento, se inicia la representación en el dominio de imágenes naturales (pesos del conjunto de datos COCO [38]), y se entrenó con 100 épocas de 1000 iteraciones cada una. Los parámetros propuestos se resumen en la Tabla 2. Se utilizaron los mismos parámetros para el entrenamiento en todos los modelos Mask R-CNN. Para entrenar el enfoque propuesto, se realizó un aumento de datos para enriquecer el conjunto de entrenamiento. Este aumento de datos consistió en una rotación de 90^{o} de todas las imágenes histológicas, con su respectiva transformación a las máscaras de entrenamiento, teniendo un total de 1282 imágenes para este

[38] Pesos del conjunto de datos COCO https://github.com/cocodataset/cocoapi

conjunto. En cuanto al enfoque, para el tercer y último nivel de representación, en el mapa de características de la última capa de la CNN Base (ResNet-50) se aplicó la estrategia SumPooling, y una posterior normalización en l2 para agrupar y fusionar las activaciones, generar nuevas regiones a partir de estas activaciones y entrenar este último modelo.

Tabla 2. Configuración de los parámetros de la propuesta.

Parámetro	Valor
Arquitectura	ResNet-50
Optimizador	SGD
Tasa de aprendizaje	0.001
Épocas	100
Iteraciones	1000
RPN Umbral NMS	0.7

5.3. Validación estadística

La validación del esquema propuesto siguió el esquema de entrenamiento-evaluación (*train-test*), de acuerdo a las particiones sugeridas por los autores del dataset evaluado. También, la hipótesis de este trabajo está determinada en la asociación de grados histológicos como segmentaciones de patrones visuales sobre las imágenes observadas. Por lo tanto, en este trabajo la validación fue enfocada principalmente en la evaluación de las segmentaciones generadas por el enfoque propuesto. A continuación se detallan las métricas utilizadas para comprender el comportamiento del enfoque propuesto.

Intersección sobre Unión (IoU): La intersección sobre unión (IoU) es una métrica que permite validar el nivel de sobrelapamiento entre la delineación realizada por un patólogo (S) y la segmentación propuesta ($\hat{S}$). En este caso, el grado de superposición entre las máscaras es definido en términos de relación de conjuntos, como:

$$IoU = \frac{S \cap \hat{S}}{S \cup \hat{S}} \tag{2}$$

Donde el numerador define el área de la intersección, y el denominador es el área de la unión entre las dos máscaras de segmentación S y $\hat{S}$. En el caso de la segmentación binaria o multi-clase, una imagen puede contener diferentes máscaras de segmentación para cada clase, por lo tanto, se calcula la media de IoU (mIoU) de una imagen tomando la IoU de cada máscara de clase segmentada, presente en la imagen original, y realizando un promedio de todas las IoU calculadas. IoU puede variar de 0 a 1, donde 0 significa que no hay superposición y 1 significa una segmentación completamente superpuesta.

Figura 17. Intersección y unión en dos máscaras de segmentación, *ground-truth* (GT) y predicha.

Área bajo la curva de la precisión y la sensibilidad (AUPRC): Esta métrica evalúa la similitud entre máscaras de segmentación teniendo en cuenta los valores de precisión y sensibilidad (recall). Estos dos conceptos tienen en cuenta los siguientes términos, que indican una contabilización de las predicciones para clasificación binaria:

- VP (Verdaderos positivos): resultados que se predijeron correctamente como positivos. En este trabajo corresponde a una segmentación estimada que corresponde con la segmentación de referencia, tomando en cuenta un valor de intersección entre las máscaras (IoU) como umbral.
- FP (Falsos positivos): resultados que se predijeron incorrectamente como positivos. En este trabajo, un FP corresponde a una máscara de segmentación estimada incorrectamente con respecto a una segmentación de referencia. En este caso, la máscara estimada tiene un nivel de IoU menor respecto a un umbral.

- VN (Verdaderos negativos): resultados que se predijeron correctamente como negativos. En este trabajo, un VN corresponde a la clase de fondo (que no contiene una máscara de segmentación) estimada con respecto a la clase de fondo de referencia, teniendo en cuenta un umbral de IoU.

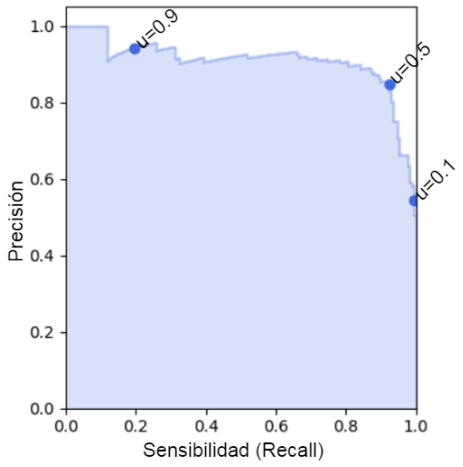
- FN (Falsos negativos): resultados que se predijeron incorrectamente como negativos. En el presente trabajo un FN corresponde a la clase de fondo estimada respecto a la de referencia, que no supera un valor respecto a un umbral de IoU.

A partir de esta configuración de estimaciones sobre las segmentaciones verdaderas y negativas, se pueden calcular estadísticas que evidencian en un mayor detalle las predicciones obtenidas. Por ejemplo, la precisión es la tasa de predicciones o clasificaciones positivas (correctas), se puede describir como: $precision = \frac{VP}{VP+FP}$. También, podemos tener en cuenta la sensibilidad (recall), referente a la tasa de positivos reales que se predijeron correctamente, y está dada por la siguiente ecuación: $recall = \frac{VP}{VP+FN}$.

Así, AUPRC compara máscaras (referencia (*ground-truth*) y predicha) calculando el valor de la precisión y la sensibilidad (recall) usando diferentes umbrales de IoU. Gráficamente se genera una curva, sensibilidad frente a precisión (la Figura 18 presenta un ejemplo), representada por cada uno de estos valores. El área bajo esta curva define el valor de la métrica AUPRC, midiendo así el balance de la precisión y la sensibilidad. Una característica de AUPRC es que no utiliza el valor de verdaderos negativos, y debido a ello esta métrica no está envuelta en gran proporción por los VN presentes en los datos, centrándose en el manejo de los ejemplos positivos, en el caso de este trabajo, los píxeles que contienen las máscaras de segmentación. Si el modelo predice correctamente los ejemplos positivos, AUPRC tendrá un valor alto. Por lo contrario, si el modelo predice incorrectamente los ejemplos positivos, el AUPRC tendrá un valor bajo.

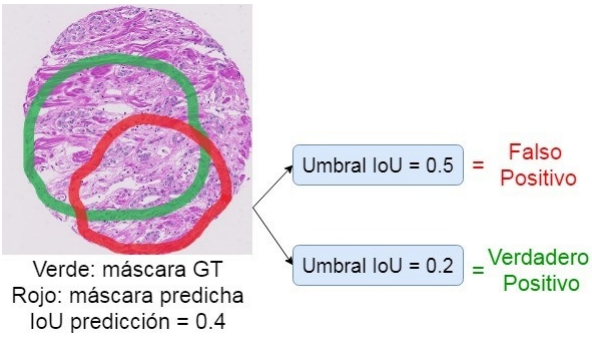
Para las tareas de segmentación de objetos, se calcula la *precision* y *recall* usando el valor de IoU para un determinado umbral. Esto significa que, para una predicción, se pueden obtener diferentes VERDADEROS o FALSOS positivos, variando el umbral de IoU (ver Figura 19). El valor de AUPRC generalmente se encuentra entre 0.5 y 1, y un valor más alto indica un

Figura 18. Ejemplo de curva precisión-sensibilidad para diferentes umbrales de probabilidad u. Los puntos marcados en color azul representan los valores de la precisión y la sensibilidad en umbrales de probabilidad 0.9, 0.5 y 0.1. El área sombreada representa la métrica AUPRC, que está dada como el área bajo la curva precisión-sensibilidad.



modelo más sensible. En este trabajo, AUPRC se calcula para cada máscara presente en la segmentación de *ground-truth* (delineación realizada por un patólogo), comparándose con la máscara de la misma clase predicha, presente en la segmentación propuesta. Ambas máscaras son máscaras binarias en donde los valores de píxeles de 0 indican que no contienen la máscara, y valores de 1 indican que contienen la máscara de segmentación.

Figura 19. Interpretación de la métrica AUPRC. En la métrica AUPRC, para distintos umbrales de IoU, se obtienen diferentes verdaderos o falsos positivos, y por tanto, diferentes valores de precisión y recuperación. En este caso, para un valor de IoU entre dos máscaras de 0.4 (en verde la máscara de *ground-truth* y en rojo la máscara predicha), debido a que es menor que un umbral de 0.5 se considera un FP. Para el mismo valor de IoU, es mayor que un umbral de 0.2 y se define como un VP.



Coeficiente de Dice: El coeficiente Dice, al igual que IoU, mide la similitud de dos muestras mediante la superposición. Esta métrica se utilizó en este trabajo como alternativa para el IoU, brindando información sobre el nivel de sobrelapamiento con mayor énfasis en los verdaderos positivos estimados en la segmentación. Esta relación entre las máscaras está definida como:

$$Dice = \frac{2|S \cap \hat{S}|}{|S| + |\hat{S}|} \tag{3}$$

El puntaje Dice por lo general tiende a ser mayor que el valor de IoU, debido a que es 2 veces el área de superposición sobre el número total de píxeles en ambas segmentaciones. Igual que IoU, un valor 0 de Dice significa que no hay superposición entre ambas máscaras y un valor de 1 significa una segmentación superpuesta de manera perfecta.

Media de la precisión promedio (mAP): Es una métrica utilizada comúnmente en enfoques de segmentación. Primero, para cada imagen, se calcula la precisión P para cada una de las máscaras de referencia (*ground-truth*) presentes, con respecto a las máscaras predichas con una misma etiqueta de clase. En este caso, se utiliza un valor de IoU para un umbral de IoU fijado: dado un umbral de IoU α. En este sentido, si IoU$>\alpha$, el valor de la precisión será 1, de lo contrario, será 0. La Tabla 3 muestra un ejemplo para el cálculo de la precisión para un umbral de IoU de 0.5. Luego de obtener la precisión para cada máscara, se calcula la precisión promedio (AP, por sus siglas en inglés) entre dos imágenes con sus máscaras de segmentación de la siguiente manera:

$$AP = \frac{\sum_{i} P(S, S^*)}{N_{mask}} \tag{4}$$

En donde P es la precisión entre la máscara delineada por un patólogo S y la máscara predicha (segmentación propuesta) $\hat{S}$, y N_{mask} es el número total de máscaras de *ground-truth* presentes en una imagen.

Por último, mAP es dado como la media de varias AP, es decir, la media de todas las precisiones

Tabla 3. Ejemplo del cálculo de la precisión y de la precisión promedio (AP) con un umbral de IoU de 0.5, entre dos imágenes con tres máscaras distintas.

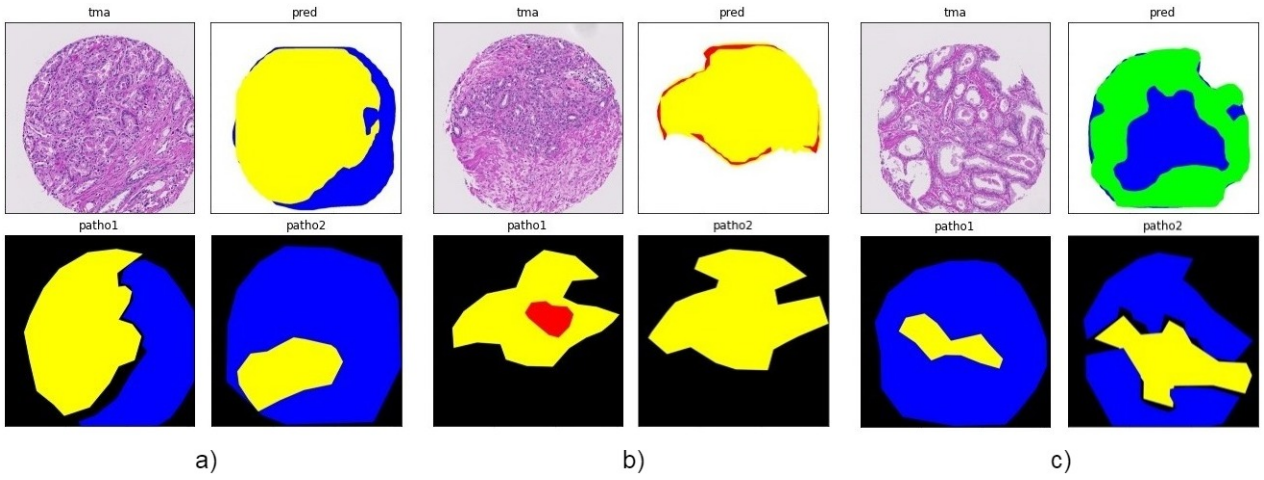
IoU	¿IoU>0.5?	P
0.3	False	0
0.6	True	1
0.7	True	1
		AP=0.666

promedio del conjunto de imágenes, y es representada como $mAP = \frac{\sum_i AP(S, S^*)}{N_{img}}$, donde AP es la precisión promedio y N_{img} es el número total de imágenes.

6. EVALUACIÓN Y RESULTADOS

La evaluación de la estrategia aquí desarrollada, se llevó a cabo para el conjunto de máscaras predichas y acopladas por los tres niveles de representación, comparado con el conjunto de anotaciones de los patólogos 1 y 2. En la Figura 20 se presentan algunos ejemplos de resultados visuales en máscaras multiclase (que poseen dos o más anotaciones de grados de Gleason) y en máscaras con una única anotación. En cada ejemplo se muestra una placa de TMA, las máscaras de segmentación predichas, generadas y acopladas por los tres niveles de representación en secuencia, y las máscaras de anotación de los patólogos 1 y 2.

Figura 20. Resultados de máscaras multiclase. Para cada ejemplo se muestra la imagen histológica, la máscara predicha por los tres niveles de representación y la delineación realizada por ambos patólogos. En cada subconjunto *tma* representa la imagen original, *pred* la imagen predicha por el método propuesto y *patho1, patho2* las anotaciones realizadas por el patólogo 1 y patólogo 2, respectivamente. En a), las máscaras poseen grados de Gleason 3 y 4 respecto a la anotación de ambos patólogos y de la predicción. En b), la máscara anotada por el patólogo 1 y la máscara predicha presentan grados 4 y 5, y la máscara anotada por el patólogo 2 presenta grado 4. En el ejemplo c), la asignación de las etiquetas por ambos patólogos fue de grados de Gleason 3 y 4, y la predicción del método presenta anotaciones con grados Benigno y Gleason 3.



La alta variabilidad descrita en los protocolos de anotación se puede ilustrar en la Figura 20-a. Las máscaras de los patólogos 1 y 2 presentan diferencias regionales en su anotación, debido a la

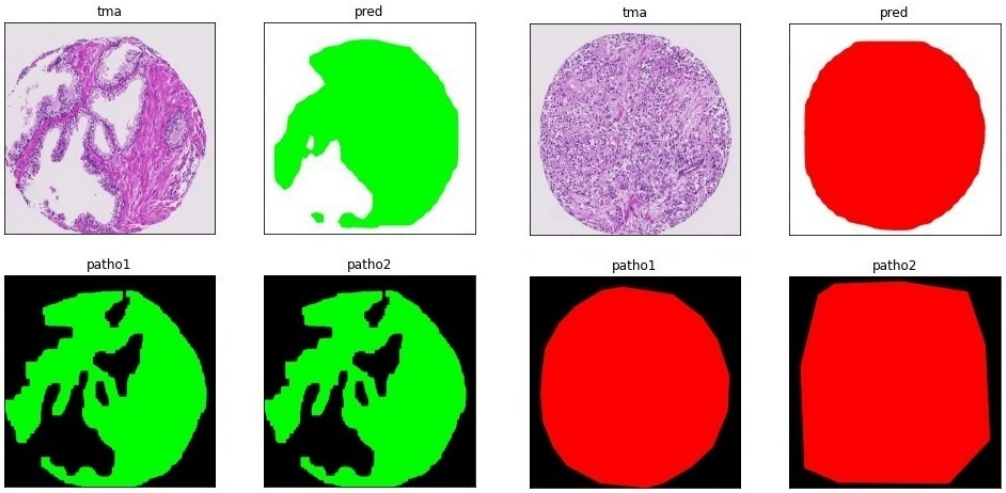
persistente subjetividad a la que está sujeto el sistema de puntuación de Gleason. Sin embargo, la máscara final predicha presenta una recuperación en las anotaciones de puntajes de Gleason 3 y 4 y una óptima segmentación comparada con la máscara del patólogo 1, y por lo tanto un aumento en las métricas a causa de los diferentes niveles de representación. En distintos casos, se presentan mejoras en la segmentación debido al tercer nivel de representación, como por ejemplo la región en amarillo (Gleason 4), que fue marcada por este último nivel, dando peso e importancia al nivel de representación Mask R-CNN implementado a partir de las activaciones de la primera red.

En la Figura 20-b, la máscara predicha presenta una recuperación en la anotación de Gleason 4 (amarillo), ya que, como se puede observar, el primer nivel de representación predijo toda la muestra como Gleason 5 (color rojo). Gracias a las nuevas anotaciones que se superponen en salidas anteriores, se presentó un aumento en las métricas de segmentación para la máscara predicha, comparada con las máscaras de anotación de ambos patólogos. Este ejemplo permite evidenciar en primera instancia la complejidad de la tarea de anotación, pero también la útilidad de comprender diferentes niveles de representación en el método propuesto, permitiendo así indagar bajo diferentes hipótesis y definiendo en conjunto una única marcación.

Como es de esperar, sin embargo, también se presentan casos en donde el método propuesto no se acerca a la delineación realizada por los patólogos, y en donde no fue capaz de caracterizar y segmentar correctamente los grados de Gleason, comparándose con las anotaciones de ambos patólogos. Un ejemplo de ello lo presenta la Figura 20-c, en donde las anotaciones realizadas por los patólogos fueron de Gleason 3 y 4, y en el caso de la predicción se caracterizaron regiones como benigno y Gleason 3. Esto puede suceder debido a la similitud estructural que existe entre grados contiguos (por ejemplo Benigno y Gleason 3), y a la complejidad de estratificar las imágenes histológicas y diferenciar sus patrones, sobre todo para grados intermedios, tres y cuatro, que presentan la principal limitación en la escala de Gleason en cuanto a su caracterización adecuada debido a su considerable similitud.

Cabe resaltar que en varias muestras la segmentación se realizó de manera óptima para máscaras

Figura 21. Resultados visuales de máscaras que poseen una única anotación de grado de Gleason. Las delineaciones realizadas por ambos patólogos y presentes en la máscara predicha se presentan en d) como Benigno (color verde) y en e) como Gleason 5 (color rojo).



de *ground truth* que contienen una única anotación de grado de Gleason, principalmente en muestras que presentan anotaciones de grados bajos (benigno y Gleason 3), como se puede observar en la Figura 21. Esto puede ser debido a que el modelo logra una buena diferenciación en los patrones glandulares para estos estadíos, a causa del tamaño y la separación de las glándulas. En los casos d) y e), la segmentación se realizó con una gran aproximación comparada con las anotaciones de ambos patólogos, y la estratificación fue realizada con éxito para los grados Benigno y Gleason 5, respectivamente.

A continuación, se presenta la evaluación cuantitativa de la estrategia desarrollada en el presente trabajo. Esta evaluación se realizó para el conjunto de máscaras predichas por los tres niveles de representación, que está notado en esta sección como Mask 3L. También, se hizo un análisis de resultados para el conjunto de máscaras generadas por el primer nivel de representación, y para el conjunto de máscaras generadas por los niveles de representación 1 y 2, y se notan como Mask 1L y Mask 2L, respectivamente. Cada conjunto fue evaluado en comparación con el subconjunto de prueba anotado por los patólogos 1 y 2. Igualmente, se calcularon las métricas entre las máscaras anotadas por los dos expertos patólogos, con el fin de obtener medidas de

referencia al momento de realizar la evaluación del enfoque propuesto en este trabajo. Las métricas se calcularon para cada clase de cada máscara de referencia (*ground-truth*). Para cada dato de prueba, el valor de una métrica estuvo dado por el promedio de los resultados por clase.

Tabla 4. Resultados de máscaras globales generadas a partir de los tres diferentes niveles de representación. Mask 1L es el conjunto de máscaras generadas por el primer modelo Mask R-CNN. Mask 2L es el conjunto de máscaras generadas a partir de la superposición de las máscaras resultantes de los modelos 1 y 2. Mask 3L es el conjunto de máscaras finales, generadas a partir de la superposición de las máscaras resultantes de los modelos 1, 2 y 3. Cada uno se comparó con las máscaras del subconjunto de prueba, para los dos diferentes patólogos. AUPRC se calculó para todo el subconjunto de prueba. Para las métricas mIoU y Dice se reportan los valores promedio y desviación estándar.

Máscaras	AUPRC	mIoU	Dice
Patólogo 1 vs Patólogo 2	0.809	0.485 ±0,326	0.574 ±0,328
Pat. 1 vs Mask 1L	0.815	0.487 ±0,391	0.536 ±0,406
Pat. 1 vs Mask 2L	**0.815**	0.496 ±0,389	0.545 ±0,405
Pat. 1 vs Mask 3L	0.805	**0.505** ±0,372	**0.566** ±0,381
Pat. 2 vs Mask 1L	0.766	0.300 ±0,339	0.351 ±0,366
Pat. 2 vs Mask 2L	0.771	0.311 ±0,341	0.364 ±0,368
Pat. 2 vs Mask 3L	0.764	0.322 ±0,327	0.385 ±0,353

Tabla 5. Resultados de mAP para cada conjunto de máscaras resultantes de los tres diferentes niveles de representación. Cada uno se comparó con las máscaras del subconjunto de prueba, para los dos diferentes patólogos.
* mAP@0.1: mAP con valor para el umbral de IoU de 0.1, mAP@0.2: mAP con valor para el umbral de IoU de 0.2, mAP@0.5: mAP con umbral de IoU de 0.5, mAP@0.7: mAP con umbral IoU de 0.7.

Máscaras	mAP@0.1	mAP@0.2	mAP@0.5	mAP@0.7
Patólogo 1 vs Patólogo 2	0.802	0.692	0.5	0.361
Pat. 1 vs Mask 1L	0.620	0.606	0.538	0.477
Pat. 1 vs Mask 2L	0.629	0.614	**0.555**	**0.485**
Pat. 1 vs Mask 3L	**0.690**	**0.659**	0.543	0.461
Pat. 2 vs Mask 1L	0.466	0.418	0.338	0.263
Pat. 2 vs Mask 2L	0.483	0.435	0.351	0.271
Pat. 2 vs Mask 3L	0.534	0.481	0.343	0.249

En la Tabla 4 se resumen los resultados obtenidos para el conjunto total de evaluación, utilizando el enfoque propuesto con diferentes niveles de representación y comparándose con respecto a cada uno de los patólogos de forma independiente. Para la validación se tuvo en cuenta las métricas de AUPRC, mIoU y Dice, que reflejan la capacidad del método propuesto en tarea regional de segmentación, con respecto a la referencia de los patólogos expertos. Como línea base para determinar el alcance del trabajo, inicialmente se decidió hacer una comparación entre las marcaciones de ambos patólogos. Como se esperaba, las marcaciones tienen una moderada coincidencia regional con niveles limitados de sobrelapamiento entre las marcaciones. Cabe destacar que el método con un mejor puntaje de AUPRC es el método propuesto Mask 2L, que integra dos redes Mask R-CNN, pero que no incluye la validación con las activaciones de la representación profunda, logrando así un 81.5 %. También en general se obtiene una asociación más cercana de las máscaras predichas con respecto al patólogo 1, esto debido a que este experto fue el mismo que hizo las anotaciones en el conjunto de datos de entrenamiento. Sin embargo, las métricas de sobrelapamiento logran un mejor desempeño utilizando tres niveles de representación (Mask 3L), lo cual puede tener ventajas en cuanto a la generalización de la propuesta. Para estas métricas, además de los valores promedio, se calcularon los valores de desviación estándar, los cuales, para todos los casos, indican una evidente variación de resultados y dispersión del conjunto de medidas de sobrelapamiento, calculadas entre máscaras de referencia y predichas.
A pesar de los desafíos inherentes para obtener una máscara ideal que determine el área de un estadío particular de Gleason, resulta relevante en este trabajo determinar la eficacia en localizar apropiadamente estos estadíos. Es por ello, que en una segunda evaluación se procedió a validar la propuesta con la metrica mAP pero sometida a diferentes umbrales del intersección de las regiones. En este caso, la métrica valida si por lo menos existe una intersección del 10 % (mAP@0.1) entre la predicción y la referencia. De este modo, puede estimar y ponderar mejor la localización que la determinación de áreas del grado de Gleason. La Tabla 5 presenta los resultados obtenidos con diferentes umbrales en la métrica mAP, siendo destacable el resultado del enfoque propuesto, utilizando la configuración completa (Mask 3L) en mAP@0.1 y

mAP@0.2, respectivamente. En este sentido se puede evidenciar que el método propuesto tiene un comportamiento apropiado para la localización de los estadíos de Gleason, pero sin tener una contundencia evidente en la segmentación de la región. También cabe resaltar que bajo estas métricas, el método propuesto alcanza considerables puntajes, comparables con los calculados entre los expertos patólogos. Lo anterior permite evidenciar el notable aporte que podría brindar esta herramienta como soporte al diagnóstico. Como se esperaba, en la validación de la localización con respecto al patólogo 2 se resalta un notable puntaje en mAP@0.1 y mAP@0.2, pero con una notable pérdida, cuando se consideran intersecciones superiores.

Asimismo, se hizo un análisis por grado de forma independiente, calculando las métricas para cada puntaje de Gleason. En este caso, para cada grado de Gleason se tomó su respectiva anotación presente en las máscaras de referencia, y se comparó con respecto a la anotación de las máscaras finales de la metodología propuesta Mask 3L (ver Tabla 6). Con esta comparación se busca indagar sobre las capacidades particulares del enfoque propuesto, con respecto a los patrones visuales que agrupan cada grado de Gleason. Como es de esperarse, el patrón aprendido que tiene una mayor asociación es el grado benigno, que resulta el más frecuente y de patrones texturales relativamente constantes en todo el conjunto de datos. De hecho, para este grado existe un puntaje significativo tanto en AUPRC como en métricas relacionadas con la segmentación regional, alcanzado puntajes de *Dice* superiores al 80 %. Este importante resultado puede sugerir la adaptabilidad de la herramienta para hacer marcaciones rápidas de tejido benigno, durante la exploración histopatológica, para que luego el experto defina regiones más específicas de otros grados afectados. En cuanto a los estadíos de Gleason 3, 4 y 5, se observa un apropiado comportamiento en términos de AUPRC, pero con notables limitaciones para seguir la delineación de los patólogos particulares que están envueltos en el presente estudio. De hecho, la mayor limitación se observa en el grado cinco, hecho que puede estar asociado por el número de muestras disponibles para el entrenamiento para esta clase, pero también la mayor variabilidad observacional en los patrones que circunscriben este nivel de la enfermedad.

En una evaluación más detallada, se realizó una evaluación discriminada por grados de Gleason

Tabla 6. Resultados entre el conjunto máscaras de los patólogos 1 y 2 por grado de Gleason, de forma independiente, y el conjunto de máscaras finales generadas por las tres redes implementadas. El número de máscaras utilizado para cada grado de Gleason anotado por cada patólogo está representado por n.

Máscaras	AUPRC	mIoU	Dice
Benigno			
Patólogo 1 vs Mask 3L (n=12)	0.908	0.747 ±0,217	0.829 ±0,214
Patólogo 2 vs Mask 3L (n=10)	0.895	0.715 ±0,224	0.804 ±0,227
Gleason 3			
Patólogo 1 vs Mask 3L (n=134)	0.758	0.545 ±0,357	0.623 ±0,361
Patólogo 2 vs Mask 3L (n=112)	0.657	0.354 ±0,285	0.459 ±0,306
Gleason 4			
Patólogo 1 vs Mask 3L (n=138)	0.772	0.365 ±0,396	0.412 ±0,426
Patólogo 2 vs Mask 3L (n=191)	0.762	0.236 ±0,333	0.281 ±0,377
Gleason 5			
Patólogo 1 vs Mask 3L (n=23)	0.685	0.252 ±0,348	0.298 ±0,380
Patólogo 2 vs Mask 3L (n=40)	0.703	0.117 ±0,259	0.141 ±0,296

Tabla 7. Resultados de mAP obtenidos para máscaras de patólogos 1 y 2 por grado de Gleason, de forma independiente, y el conjunto de máscaras finales generadas por las tres redes implementadas. El número de máscaras utilizado para cada grado de Gleason anotado por cada patólogo está representado por n.

Máscaras	mAP@0.1	mAP@0.2	mAP@0.5	mAP@0.7
Benigno				
Patólogo 1 vs Mask 3L (n=12)	0.916	0.916	0.916	0.916
Patólogo 2 vs Mask 3L (n=10)	0.9	0.9	0.9	0.9
Gleason 3				
Patólogo 1 vs Mask 3L (n=134)	0.799	0.739	0.567	0.493
Patólogo 2 vs Mask 3L (n=112)	0.786	0.589	0.313	0.170
Gleason 4				
Patólogo 1 vs Mask 3L (n=138)	0.514	0.486	0.399	0.283
Patólogo 2 vs Mask 3L (n=191)	0.372	0.361	0.257	0.173
Gleason 5				
Patólogo 1 vs Mask 3L (n=23)	0.435	0.391	0.217	0.174
Patólogo 2 vs Mask 3L (n=40)	0.2	0.2	0.1	0.1

y teniendo en cuenta diferentes niveles de umbral de intersección de regiones, en la cuantificación del mAP. En la Tabla 7 se reportan los resultados para los diferentes niveles de Gleason. Como es de esperarse, el nivel benigno, con patrones planos y poca variabilidad tienen una alta corres-

pondencia tanto en la localización (mAP@0.1), como en puntajes asociados con la segmentación (mAP@0.7). También es notable el desempeño logrado en Gleason 3, sobre todo en términos de localización del estadío. Los demás grados presentan notables dificultades y su caracterización puede requerir muchas más anotaciones, así como también la delineación por parte de múltiples expertos. En consecuencia, el trabajo presentado tiene notables ventajas como herramienta para el soporte clínico, pero acotado a una intervención posterior por parte del patólogo para tomar decisiones en las delineaciones finales de los grados de Gleason. Por ejemplo, en estadíos tempranos y para tejido benigno, el enfoque propuesto puede mostrar considerables ventajas para hacer la tarea automática, agilizando el proceso de análisis y permitiendo a los expertos enfocarse en regiones de las placas más críticas. Por otra parte, la visualización de características durante la delineación, en el tercer nivel de la representación propuesta, puede apoyar la tarea del patólogo y reducir el sesgo en las anotaciones.

7. CONCLUSIONES Y PERSPECTIVAS

En el presente trabajo se introdujo una representación profunda dedicada a la segmentación de patrones visuales con correspondencia a los estadíos de Gleason, asociados a la agresividad del cáncer de próstata en imágenes histológicas. La representación propuesta se fundamenta en la arquitectura Mask R-CNN que ha demostrado notables resultados en la segmentación semántica de instancias naturales. Sin embargo, debido a la complejidad de la tarea histopatológica, esta representación resulta insuficiente utilizando una única etapa de aprendizaje. Es por ello que en este trabajo se introduce un esquema con diferentes niveles de representación, basado en la totalidad de datos de entrenamiento (primer nivel), utilizando las regiones más desafiantes para el modelo, es decir, aquellas regiones en las que el primer nivel predecía dos grados distintos de Gleason (segundo nivel), y definiendo un esquema basado en las regiones de mayor atención visual (tercer nivel). Esta estrategia de refinamiento conduce a una mejora en el desempeño, en lo que se refiere al nivel de sobrelapamiento entre máscaras de referencia y predichas, evidenciada en los valores resultantes de las métricas mIoU y Dice. El último nivel, además, pretende reducir un poco el sesgo marcado por las anotaciones del patólogo.

El enfoque propuesto mostró resultados relevantes en cuanto a la localización de los estadíos de Gleason, principalmente para los grados Benigno y Gleason 3, a partir del método de segmentación Mask R-CNN, que permite lograr un aprendizaje de características para la determinación y delimitación de regiones con significado. Esta herramienta entonces permite estratificar y segmentar muestras histológicas a partir de patrones globales, y patrones regionales o locales que permiten sugerir y realizar nuevas marcaciones sobre máscaras anteriormente generadas. A partir de estas representaciones, se calcularon diferentes máscaras de segmentación en secuencia que representan regiones asociadas al cáncer de próstata en imágenes histológicas. En el trabajo se consideró un problema multiclase, entrenando la representación con cuatro posibles estadíos del cáncer. Los modelos de este método, abordados en la estrategia y acoplados en secuencia, presentan resultados experimentales similares a los resultados interpatólogos, especialmente

comparándose con el patólogo 1, y se demuestra que las redes convolucionales profundas pudieron ser entrenadas con éxito como anotador de puntaje de Gleason, con una medida de sobrelapamiento (mIoU) de 50.5 % teniendo en cuenta los tres niveles de representación (Mask 3L) con respecto al patólogo 1, valor que supera el mIoU calculado entre patólogos.

En cuanto al análisis y estratificación de los estadíos del cáncer, una de las limitaciones inherentes es la moderada concordancia entre expertos patólogos para definir las fronteras de afectación y los estadíos asociados en cada nivel. Este hecho ha sido ampliamente fundamentado en los diversos estudios de concordancia y variabilidad en la delimitación de regiones y diagnóstico de placas histopatológicas. Para soportar esta tarea, previamente se han propuesto diversos enfoques, basados en la extracción de características en imágenes histológicas para su posterior clasificación. Sin embargo, estos enfoques se limitan a un número de características predefinidas y, por consiguiente, no generalizan completamente la enfermedad. Otros enfoques se centran en la extracción de parches histológicos para su posterior clasificación. No obstante, estos enfoques son limitados a una caracterización local de regiones sin tener muchas veces en cuenta las estructuras celulares que se correlacionan con la enfermedad. Se concluye que, en contraste, el enfoque propuesto utiliza regiones completas, con sentido histopatológico, que abarca diferentes estructuras celulares para realizar la aproximación en la anotación automática. En este sentido, la herramienta propuesta puede ser importante para proponer candidatos de regiones con un estadío de Gleason particular, el cual puede luego ser modificado más fácilmente por un experto. En este sentido, también se han propuesto algunas aproximaciones que buscan segmentar glándulas y estructuras celulares específicas. Esta tarea resulta interesante para apoyar la tarea diagnóstica, pero los datos específicos de entrenamiento hacen la tarea tediosa, con conjuntos de datos limitados, siendo una tarea compleja el incremento dinámico del conjunto de datos. Estas herramientas también, al limitarse en las estructuras celulares conocidas, impiden explorar nuevas relaciones o patrones arquitecturales de la célula, que puedan ser auto-aprendidos por los algoritmos y puedan definir un estadío de Gleason particular.

El presente trabajo tiene un amplio potencial para ser implementado como herramienta de so-

porte en etapas preliminares para proponer regiones y grados de Gleason, que posteriormente sean ajustados y validados por patólogos. Sin embargo, el presente enfoque tiene una amplia dependencia de las anotaciones aprendidas por el patólogo 1, lo cual lo hace susceptible a delimitaciones fijas sobre patrones comunes marcados por el experto, los cuales pueden distar ampliamente del patólogo 2, como se reportaron en la sección de resultados. Trabajos futuros incluyen la validación con el mismo conjunto de datos, pero anotado por un mayor número de patólogos, permitiendo introducir una mayor flexibilidad en la representación aprendida. A partir de ello, se espera formular estrategias que ajusten las diferentes anotaciones, con respecto a la experticia de cada patólogo involucrado en el estudio. También se realizarán pruebas del enfoque con grados de Gleason únicos, que permitan enfocarse en patrones específicos y puedan ser herramientas más ligeras y con mayor probabilidad de acierto para proponer regiones candidatas de manera efectiva.

8. ANEXOS

8.1. Productos

Artículo de investigación

- Andrés Gómez, Fabian León, Miguel Plazas y Fabio Martínez. "Segmentación multinivel de patrones de Gleason en imágenes histopatológicas". Sometido a Revista TecnoLógicas (2021).

BIBLIOGRAFÍA

American Cancer Society. *Cómo entender su informe de patología: cáncer de próstata*. En: cancer.org. 2017 (vid. págs. 15, 17).

— *Pruebas para diagnosticar y determinar la etapa del cáncer de próstata*. En: cancer.org. 2019 (vid. págs. 12, 16, 17).

American Society of Clinical Oncology. *Cáncer de próstata: Estadísticas*. En: Cancer.Net. 2018 (vid. pág. 28).

Arvaniti, Eirini y col. "Automated Gleason grading of prostate cancer tissue microarrays via deep learning". En: *Scientific reports* (2018) (vid. págs. 12, 19).

— "Replication Data for: Automated Gleason grading of prostate cancer tissue microarrays via deep learning". En: (2018) (vid. pág. 43).

Babenko, A. y V. Lempitsky. "Aggregating local deep features for image retrieval". En: *International Conference on Computer Vision (ICCV)* (2015) (vid. pág. 40).

Bley, Enrique y Andrés Silva. "Diagnóstico precoz del cáncer de próstata". En: *Revista médica clínica Las Condes* 22.4 (2011), págs. 453-458 (vid. pág. 11).

Bray, F. y col. "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(6) (2018), págs. 394-424 (vid. pág. 11).

Bulten, Wouter y col. "Automated gleason grading of prostate biopsies using deep learning". En: *arXiv preprint arXiv:1907.07980* (2019) (vid. pág. 19).

Comisión de Salud Pública. “Evaluación del Sistema Gleason”. En: *Urología Colombiana* IX.1 () (vid. págs. 12, 17).

D F R Griffiths y col. “A study of Gleason score interpretation in different groups of UK pathologists; techniques for improving reproducibility”. En: *Histopathology* 48.6 (2006), págs. 655-662 (vid. págs. 12, 17).

Dai, Jifeng y col. “R-FCN: Object Detection via Region-based Fully Convolutional Networks”. En: *arXiv:1605.06409v2* (2016) (vid. pág. 24).

Doyle, S. y col. “Automated Grading of Prostate Cancer using Architectural and Textural Image Features”. En: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2007), págs. 1284-1287 (vid. págs. 12, 18).

Doyle, Scott y col. “A Boosting Cascade for Automated Detection of Prostate Cancer from Digitized Histology”. En: *Dept. of Biomedical Engineering, Rutgers Univ., Piscataway, NJ 08854, USA, Dept. of Surgical Pathology, Univ. of Pennsylvania, Philadelphia, PA 19104, USA* IX.1 () (vid. págs. 13, 18).

Farjam, R. y col. “An Image Analysis Approach for Automatic Malignancy Determination of Prostate Pathological Images”. En: *Cytometry. Part B, Clinical cytometry* 72 (2007), págs. 227-40 (vid. pág. 18).

García, José Gabriel, Valery Naranjo y Adrián Colomer. “Diseño y desarrollo de un sistema automático de clasificación de estructuras glandulares en imágenes histológicas de próstata”. En: *Universidad Politécnica de Valencia* (2018) (vid. pág. 13).

Girshick, Ross. “Fast R-CNN”. En: *arXiv:1504.08083v2* (2015) (vid. pág. 21).

Girshick, Ross y col. “Region-based Convolutional Networks for Accurate Object Detection and Segmentation”. En: *UC Berkeley* () (vid. pág. 20).

Gurcan, Metin N y col. “Histopathological image analysis: A review”. En: *IEEE reviews in biomedical engineering* 2 (2009), págs. 147-171 (vid. pág. 18).

He, Kaiming y col. “Mask R-CNN”. En: *arXiv:1703.06870v3* (2018) (vid. pág. 33).

Ing, Nathan y col. “Semantic segmentation for prostate cancer grading by convolutional neural networks”. En: *ResearchGate* (2018) (vid. pág. 19).

Instituto Nacional del Cáncer. *Tinción con hematoxilina y eosina* (vid. pág. 15).

Kalantidis, Y., C. Mellina y S. Osindero. “Crossdimensional weighting for aggregated deep convolutional features”. En: *arXiv:1512.04065* (2015) (vid. pág. 40).

LeCun, Y., Bengio Y. e Hinton, G. “Deep learning”. En: *Nature* 521 (2015), 436–444 (vid. pág. 20).

León, Fabian, Miguel Plazas y Fabio Martínez. “An inception deep architecture to differentiate close-related Gleason prostate cancer scores”. En: (2019) (vid. págs. 13, 19).

Lin, Tsung-Yi y col. “Focal Loss for Dense Object Detection”. En: *arXiv:1708.02002v2* (2018) (vid. pág. 25).

M McLean y col. “Interobserver variation in prostate cancer gleason scoring: are there implications for the design of clinical trials and treatment strategies?” En: *Clinical oncology* 9.4 (1997), págs. 222-225 (vid. págs. 12, 17).

Morera, Pamela Bolaños y Carolina Chacón Araya. "Escala patológica de Gleason para el cáncer de prostata y sus modificaciones". En: *Medicina Legal de Costa Rica* 34.1 (2017) (vid. pág. 16).

Nassar, Aziza. "Biopsia: 5 cosas que todo paciente debe saber". En: *Cancer.Net* (2017) (vid. pág. 15).

Payá, Elena, Valery Naranjo y Jose García. "Diseño y desarrollo de un sistema automático de segmentación de glándulas histológicas para identificar el cáncer de próstata en una etapa inicial". En: *Universidad Politécnica de Valencia* (2019), pág. 5 (vid. págs. 13, 16).

Ramos, Christian, Juan Fullá y Alejandro Mercado. "Detección precoz de cáncer de próstata: Controversias y recomendaciones actuales". En: *Revista médica clínica Las Condes* 29.2 (2018), págs. 128-135 (vid. pág. 15).

Redmond, J. y col. "You Only Look Once: Unified, Real-TimeObject Detection". En: *arXiv:1506.02640v5* (2016) (vid. pág. 27).

Ren, Shaoqing y col. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". En: *arXiv:1506.01497v3* (2016) (vid. pág. 22).

Ruiz, Ana Isabel y col. "Actualización sobre cáncer de próstata". En: *Correo Científico Médico* 21.3 (2017) (vid. págs. 11, 15).

Tong, Simon y Daphne Koller. "Support Vector Machine Active Learning with Applications to Text Classification". En: *Journal of Machine Learning Research* (2001) (vid. pág. 21).

Wu, Jianxin. "Convolutional neural networks". En: *National Key Lab for Novel Software Technology Nanjing University, China* (2020) (vid. pág. 20).

Zhong, Qing y col. “A curated collection of tissue microarray images and clinical outcome data of prostate cancer patients”. En: *Sci Data 4, 170014* (2017) (vid. pág. 43).