

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería y Ciencias
Maestría en Ciencia de Datos
Proyecto Aplicado III

Diseño e implementación de modelos de Aprendizaje de
Instancias Múltiples para la clasificación débilmente supervisada
de imágenes histopatológicas de cáncer de próstata

Juan José Restrepo Rosero
María Valentina Belalcázar Perdomo

Director: Dr. Julián Gil González

24 de septiembre de 2025



Nota de Aceptación

Aprobado por el Comité de Trabajo de Grado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana para optar al título de Magíster en Ciencia de Datos.

Dr. Diego Luis Linares Ospina
Director Maestría en Ciencia de Datos

Dr. Julián Gil González
Director Trabajo de Grado

Dr. Nombre Apellido
Jurado

Dr. Nombre Apellido
Jurado

Santiago de Cali, 24 de septiembre de 2025

Señores
Pontificia Universidad Javeriana – Cali

Decano
Facultad de Ingeniería y Ciencias
Cali

Cordial Saludo.

Nos permitimos presentar a su consideración el Proyecto Aplicado titulado “Diseño e implementación de modelos de Aprendizaje de Instancias Múltiples para la clasificación débilmente supervisada de imágenes histopatológicas de cáncer de próstata” con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el proyecto de grado y posteriormente optar al título de Magíster en Ciencia de Datos.

Al firmar aquí, damos fe que entendemos y conocemos las directrices para la presentación de trabajos de grado de la Facultad de Ingeniería y Ciencias aprobadas el 26 de noviembre de 2009, donde se establecen los plazos y normas para el desarrollo del anteproyecto y del trabajo de grado.

Atentamente,

Juan José Restrepo R.

Juan José Restrepo Rosero
Código: 8939280



María Valentina Belalcázar Perdomo
Código: 8992284

Santiago de Cali, 24 de septiembre de 2025

Señores
Pontificia Universidad Javeriana – Cali

Decano
Facultad de Ingeniería y Ciencias
Cali

Cordial Saludo.

Por medio de la presente me permito informarle que los estudiantes de la Maestría en Ciencia de Datos Juan José Restrepo Rosero (cod: 8939280) y María Valentina Belalcázar Perdomo (cod: 8992284) trabajaron bajo mi dirección en el Proyecto Aplicado titulado ‘Diseño e implementación de modelos de Aprendizaje de Instancias Múltiples para la clasificación débilmente supervisada de imágenes histopatológicas de cáncer de próstata’.

Atentamente,

julián gil
gonzález

Firmado digitalmente
por julián gil gonzález
Fecha: 2025.09.29
08:56:52 -05'00'

Dr. Julián Gil González
Director Trabajo de Grado

Resumen

El presente proyecto propone el diseño e implementación de un pipeline reproducible basado en técnicas de Aprendizaje de Instancias Múltiples (Multiple Instance Learning, MIL) para la clasificación débilmente supervisada de Whole Slide Images (WSI) de cáncer de próstata. Frente a la elevada variabilidad interobservador en la evaluación histológica y la escasez de anotaciones a nivel de parche, el trabajo investigará variantes de MIL que integren extractores de características robustos, mecanismos de atención y esquemas de agregación multinivel, además de enfoques probabilísticos para estimar incertidumbre y mejorar la explicabilidad. La metodología combinará preprocesamiento y extracción de parches, extracción de embeddings por redes convolucionales y módulos de agregación entrenados con etiquetas a nivel de lámina; la evaluación se realizará sobre la base de datos pública SICAPv2 empleando validación cruzada estratificada por paciente y métricas de evaluación (AUC, precisión, recall, F1, coeficiente Kappa), complementadas con análisis de incertidumbre. Se espera entregar datasets y pipelines versionados, implementaciones reproducibles de las variantes MIL, un estudio comparativo riguroso y salidas explicables que faciliten la validación clínica.

Palabras Claves: Aprendizaje de instancias múltiples; Imágenes de diapositivas completas; Redes neuronales convolucionales; Histopatología; Cáncer de próstata

Índice general

| | |
|---|-----------|
| 1. Introducción | 1 |
| 2. Planteamiento del Problema | 2 |
| 2.1. Planteamiento del problema..... | 2 |
| 3. Formulación del Problema | 4 |
| 3.1. Pregunta general..... | 4 |
| 3.2. Preguntas de sistematización | 4 |
| 4. Objetivos | 5 |
| 4.1. Objetivo General | 5 |
| 4.2. Objetivos Específicos..... | 5 |
| 5. Resultados Esperados | 6 |
| 6. Alcances y Limitaciones | 7 |
| 6.1. Alcances | 7 |
| 6.2. Limitaciones | 7 |
| 7. Justificación | 9 |
| 7.1. Justificación..... | 9 |
| 8. Marco Teórico | 11 |
| 8.1. Glosario | 11 |
| 8.2. Marco Conceptual..... | 11 |
| 8.2.1. Cáncer de próstata y escala de Gleason | 11 |
| 8.2.2. Histopatología digital y Whole Slide Images (WSI)..... | 12 |
| 8.2.3. Redes neuronales convolucionales como extractores de características | 13 |
| 8.2.4. Aprendizaje de Instancias Múltiples (MIL) | 13 |
| 8.2.5. Métricas de evaluación e interpretabilidad | 14 |
| 9. Marco de Antecedentes | 15 |
| 9.0.1. CAMIL: channel attention-based multiple instance learning for whole slide image classification..... | 15 |
| 9.0.2. A Probabilistic MIL Model with Spatial Regularization for Weakly Supervised Histopathology Image Analysis..... | 16 |
| 9.0.3. Graph-based Multiple Instance Learning for Whole Slide Image Classification | 16 |
| 9.0.4. Deep Recurrent Attention MIL for Histopathological Image Analysis..... | 16 |
| 9.0.5. Hierarchical Pooling in MIL for Histopathological Subtype Classification | 17 |
| 10. Metodología | 18 |

| | |
|--|-----------|
| 11. Recursos | 20 |
| 11.1. Recursos | 20 |
| 11.1.1. Participantes..... | 20 |
| 11.1.2. Descripción general de recursos materiales y su disponibilidad | 20 |
| 11.1.3. Aseguramiento de acceso y validación de disponibilidad | 21 |
| 11.1.4. Entregables relacionados con recursos | 21 |
| 12. Cronograma | 22 |
| 12.0.1. Cronograma | 22 |
| Bibliografía | 23 |

Capítulo 1

Introducción

Las técnicas de diagnóstico asistido por inteligencia artificial en histopatología han mostrado un notable potencial para aumentar la precisión, la reproducibilidad y la velocidad del análisis de tejido, contribuyendo a reducir la variabilidad interobservador y los tiempos de informe clínico [1] [2]. No obstante, en patologías como el cáncer de próstata las Whole Slide Images (WSI) suelen estar etiquetadas a nivel de muestra (slide-level) y no parche a parche (instancia), lo que dificulta el uso directo de enfoques totalmente supervisados que requieren anotaciones finas [3]. Frente a esta limitación, el paradigma de aprendizaje de instancias múltiples (Multiple Instance Learning, MIL) surge como una alternativa práctica: permite entrenar modelos empleando etiquetas globales por WSI, reduciendo la carga de anotación experta sin renunciar a rendimiento clínico [4] [5].

El presente proyecto propone diseñar un pipeline modular y reproducible para aplicar MIL a WSI de cáncer de próstata (dataset SICAPv2). El pipeline integrará etapas de preprocesamiento y extracción de parches, extracción de embeddings por redes convolucionales, y módulos de agregación débilmente supervisados que incorporen explicabilidad y estimación de incertidumbre, con especial atención a la reproducibilidad experimental [6] [7] [8]. La evaluación seguirá protocolos clínicamente relevantes (validación cruzada estratificada por paciente) y reportará métricas estándar (AUC, precisión, recall, F1) junto con coeficientes de concordancia (Kappa) y análisis estadísticos que permitan comparar métodos más allá de cifras puntuales.

Como resultado de esta investigación, se espera obtener una serie de entregables concretos y de alto valor. En primer lugar, se generará un conjunto de parches reproducible y versionado, derivado del dataset SICAPv2, que incluirá metadatos detallados y splits organizados por paciente para garantizar la validez de las pruebas. Paralelamente, se pondrán a disposición implementaciones de código abierto de diversas variantes de MIL y extractores de características, junto con sus correspondientes checkpoints y scripts de entrenamiento y evaluación. Finalmente, y como culminación del proyecto, se presentará un estudio comparativo riguroso que no solo documentará los trade-offs entre precisión, capacidad de generalización y explicabilidad, sino que también estará apoyado en estimaciones de incertidumbre y protocolos de validación clínica. Estos entregables, en su conjunto, buscan aportar evidencia técnica y clínica sólida que justifique la integración de herramientas de apoyo al diagnóstico, facilitando así la interpretación y la eventual adopción por parte de los especialistas.

Capítulo 2

Planteamiento del Problema

2.1. Planteamiento del problema

El cáncer de próstata constituye una de las principales causas de morbilidad y mortalidad en la población masculina a nivel mundial. Según estimaciones recientes de la Organización Mundial de la Salud (OMS) y del Global Cancer Observatory (GCO), se reportan anualmente más de 1.4 millones de casos nuevos, con una carga especialmente elevada en países de ingresos medios y bajos por las barreras que afectan el acceso al diagnóstico temprano y al tratamiento oportuno [9]. En el contexto colombiano, esta enfermedad figura entre las primeras causas de muerte por cáncer en hombres, lo que subraya la urgencia de mejorar la detección precoz y la estratificación del riesgo para reducir la morbimortalidad asociada [10].

El diagnóstico histopatológico mediante la escala de Gleason es el pilar de la estratificación de riesgo; no obstante, su aplicación presenta variabilidad interobservador. Esta subjetividad es especialmente relevante en la discriminación entre patrones limítrofes (por ejemplo, Gleason 3 frente a Gleason 4) y en la identificación de subtipos de mal pronóstico como el patrón *cribriforme*. El patrón cribiforme se caracteriza por formaciones glandulares con arquitectura tipo “criba”, o agregados celulares que generan múltiples lúmenes o masas con espacios intercalados; su presencia se asocia a mayor agresividad y riesgo de recurrencia, por lo que su correcta detección tiene implicaciones terapéuticas relevantes [11] [12].

La digitalización de muestras y la generación de Whole Slide Images (WSI) habilitan el análisis computacional del tejido completo, pero introducen retos técnicos importantes: tamaño gigapíxel de las imágenes, heterogeneidad morfológica, variaciones de tinción entre centros, y artefactos de preparación y escaneo [13]. Además, disponer de conjuntos con suficiente cantidad de muestras etiquetadas a nivel de parche es costoso y poco frecuente; en la práctica muchas bases de datos ofrecen únicamente etiquetas a nivel de la lámina completa.

Es importante aclarar la diferencia de etiquetas que condiciona la metodología: una etiqueta a nivel de muestra(slide-level) asigna una única etiqueta a toda la WSI (por ejemplo, presencia/ausencia de tumor o puntaje Gleason global), mientras que una etiqueta a nivel de instancia (patch-level) identifica la etiqueta de cada parche individual extraído de la WSI. La ausencia generalizada de anotaciones patch-level explica el interés por métodos de aprendizaje débil como Multiple Instance Learning (MIL), que entran modelos usando solo etiquetas slide-level.

Desde la perspectiva de la ciencia de datos, la combinación de estas condiciones genera un conjunto de desafíos metodológicos y experimentales bien definidos. Por un lado, el manejo eficaz de datos masivos exige estrategias de preprocesamiento, muestreo y representación que preserven información diagnóstica relevante sin sobrecargar la infraestructura de cómputo. Por otro lado, la escasez y el carácter generalmente ‘débil’ de las etiquetas, es decir, se dispone únicamente de etiquetas globales asignadas a toda la lámina histológica (WSI), sin anotaciones detalladas para cada parche o región de tejido, plantea problemas de aprendizaje con

supervisión limitada, desbalance de clases y evaluación fiable, especialmente cuando los subtipos de interés son poco prevalentes. Además, la heterogeneidad entre protocolos de tinción y escaneo introduce *shifts* de dominio que comprometen la generalización interinstitucional de los modelos, y la probable presencia de artefactos exige procedimientos de control de calidad y métricas de robustez [2] [13].

En cuanto a la trazabilidad experimental, el proyecto adoptará prácticas esenciales de reproducibilidad, por ejemplo, particiones estratificadas por paciente, registro de semillas y versionado del conjunto de parches para garantizar comparaciones internas consistentes entre métodos. Otras prácticas más extensas de validación interinstitucional o evaluaciones cualitativas amplias con múltiples patólogos se consideran relevantes, pero exceden el alcance temporal de esta fase y quedan planteadas como trabajo futuro.

En síntesis, el problema que aborda este proyecto es aplicar y adaptar métodos de aprendizaje débil (MIL) y estrategias de preprocesamiento para construir un pipeline práctico y evaluable capaz de clasificar WSI prostáticos con etiquetas slide-level, enfocándose en la validez operativa y la evaluación cuantitativa del rendimiento en conjuntos con disponibilidad limitada de anotaciones finas.

Capítulo 3

Formulación del Problema

3.1. Pregunta general

Teniendo en cuenta lo anterior y las promesas de las arquitecturas MIL para el análisis de imágenes histopatológicas, surge la siguiente pregunta de investigación:

¿Cómo diseñar, implementar y evaluar un modelo de visión por computador basado en técnicas de Aprendizaje de Instancias Múltiples (MIL), que usando únicamente etiquetas a nivel de Whole Slide Image (WSI) maximice la precisión diagnóstica para la clasificación de cáncer de próstata?

3.2. Preguntas de sistematización

Para responder a la pregunta general se plantean las siguientes preguntas de sistematización, que dividen el problema en subretos concretos y evaluables:

1. ¿Qué técnicas de preprocessamiento y extracción de parches (normalización de tinción, filtrado por tejido, selección de resolución y aumentos específicos) optimizan la calidad de las instancias de entrada para tareas de clasificación en WSI prostáticos?
2. ¿Qué estrategias de extracción de características y agregación en el paradigma MIL (uso de extractores CNN / transfer learning y distintos esquemas de fusión) permiten integrar evidencia local y contexto global para mejorar la clasificación de patrones Gleason y subtipos relevantes?
3. ¿Cómo evaluar de manera rigurosa el rendimiento y la incertidumbre de las arquitecturas propuestas mediante protocolos reproducibles (validación cruzada estratificada por paciente) y métricas estándar (AUC, precisión, recall, F1-score y coeficiente Kappa), de modo que permitan comparaciones estadísticas y decisiones clínicamente relevantes?

Capítulo 4

Objetivos

4.1. Objetivo General

Desarrollar un modelo de visión por computador a partir de técnicas de Aprendizaje con Múltiples Instancias para la clasificación débilmente supervisada de imágenes de histopatología de cáncer de próstata, con el fin de maximizar la precisión diagnóstica y facilitar su interpretación clínica.

4.2. Objetivos Específicos

1. Implementar técnicas de procesamiento de imágenes de histopatología prostática, incluyendo corrección de color, eliminación de artefactos y normalización de tinciones.
2. Implementar modelos de visión por computador basados en Aprendizaje de Instancias Múltiples con Deep Learning para la clasificación de Whole Slide Images.
3. Evaluar el rendimiento de los modelos desarrollados mediante métricas estándar (AUC, precisión, recall, F1-score y coeficiente Kappa) con validación cruzada.

Capítulo 5

Resultados Esperados

A medida que se cumple con la metodología establecida, se espera obtener los siguientes resultados, para dar como culminado el desarrollo del proyecto:

1. Un conjunto de datos de parches histopatológicos curado y versionado, derivado de la base SICAPv2 y optimizado mediante técnicas de preprocesamiento clave para el diagnóstico de cáncer de próstata.
2. Un conjunto de modelos de Aprendizaje de Instancias Múltiples (MIL) basados en arquitecturas de deep learning, incluyendo mecanismos de atención, pooling jerárquico y variantes avanzadas como Graph-MIL y VGPMIL-PR, adaptadas para la clasificación de imágenes histopatológicas prostáticas.
3. Un estudio comparativo de desempeño de modelos, documentado con métricas cuantitativas (AUC, precisión, recall, F1-score y coeficiente Kappa de Cohen), empleando validación cruzada estratificada por paciente, con el fin de seleccionar la arquitectura que ofrezca el mejor equilibrio entre precisión diagnóstica y capacidad de generalización.
4. Una documentación técnica y científica completa del pipeline, abarcando desde el preprocesamiento hasta la interpretación de salidas, estructurada para servir como referencia en investigaciones futuras y potencialmente adaptable a otros tipos de cáncer o patologías histopatológicas.

Capítulo 6

Alcances y Limitaciones

6.1. Alcances

El proyecto abarcará el desarrollo, implementación y evaluación de un modelo de visión por computador basado en aprendizaje con múltiples instancias (Multiple Instance Learning, MIL) para la clasificación débilmente supervisada de imágenes histopatológicas (Whole Slide Images, WSI) de cáncer de próstata, utilizando como fuente de datos la base pública SICAPv2.

El trabajo incluirá:

- El preprocesamiento y segmentación de las WSI en parches optimizados para entrenamiento.
- El diseño e implementación de distintas arquitecturas MIL (atención, pooling jerárquico, Graph-MIL, VGPMIL-PR) en entornos Python con PyTorch y/o TensorFlow.
- La evaluación de los modelos mediante métricas estándar (AUC, precisión, recall, F1-score y coeficiente Kappa de Cohen) y validación cruzada estratificada por paciente.
- La incorporación de mecanismos básicos de explicabilidad, como mapas de atención, para interpretar las decisiones del modelo

No se contempla en este proyecto:

- La validación clínica del modelo en entornos hospitalarios reales.
- La integración del sistema en plataformas de diagnóstico médico.
- El desarrollo de una interfaz de usuario o software listo para uso clínico.
- La inclusión de variables clínicas adicionales (Por ejemplo, niveles de PSA, historial médico, entre otras).

Estos aspectos podrán ser abordados en trabajos futuros para ampliar la aplicabilidad del modelo.

6.2. Limitaciones

- **Base de datos:** El entrenamiento y evaluación se realizarán exclusivamente con SICAPv2, lo que podría limitar la generalización a imágenes provenientes de otras instituciones o con diferentes protocolos de escaneo.
- **Etiquetado:** El modelo se entrenará únicamente con etiquetas globales a nivel de WSI (bags), sin anotaciones por parche, lo que podría afectar la precisión en la detección de patrones histológicos muy localizados.

- **Recursos computacionales:** El procesamiento de WSI y el entrenamiento de modelos MIL requieren hardware especializado (GPU de alto rendimiento), por lo que la optimización del código será esencial para garantizar tiempos de ejecución viables.
- **Cobertura diagnóstica:** El proyecto se centrará exclusivamente en cáncer de próstata, sin abordar la clasificación de otros tipos de cáncer o lesiones benignas.

Estas limitaciones no comprometen la viabilidad ni la pertinencia del proyecto, ya que el enfoque propuesto está alineado con las mejores prácticas de investigación en histopatología computacional y permitirá establecer una base sólida para futuras mejoras y aplicaciones clínicas.

Capítulo 7

Justificación

7.1. Justificación

La mejora en la precisión y eficiencia del diagnóstico del cáncer de próstata representa una necesidad prioritaria en el campo de la salud pública. Esta enfermedad continúa siendo una de las principales causas de mortalidad en hombres a nivel global y nacional, con impactos directos en la calidad de vida de los pacientes y en la carga del sistema sanitario. El método diagnóstico estándar (la evaluación histopatológica de biopsias) depende en gran medida de la interpretación visual del patólogo, lo que introduce un alto grado de variabilidad interobservador, especialmente en la clasificación de patrones intermedios como el Gleason 4 y subtipos agresivos como el cribiforme, asociados a peores pronósticos clínicos [14].

En este contexto, el uso de técnicas de aprendizaje automático, y en particular del paradigma de aprendizaje de instancias múltiples (Multiple Instance Learning, MIL), ofrece una solución viable para automatizar el análisis de imágenes histológicas de Whole Slide Images (WSI), caracterizadas por su gran tamaño y complejidad estructural. A diferencia del aprendizaje supervisado convencional, que exige anotaciones exhaustivas a nivel de parche, el enfoque MIL permite entrenar modelos utilizando únicamente etiquetas globales por imagen, reduciendo drásticamente la necesidad de recursos humanos especializados para el etiquetado y aumentando la escalabilidad de los sistemas [2] [4].

La viabilidad técnica de este proyecto se sustenta en el uso de la base de datos pública y validada SICAPv2, la cual contiene 155 WSI de cáncer de próstata correspondientes a 95 pacientes, junto con un total de 18,783 parches etiquetados de tamaño 512×512 píxeles a una ampliación de 10×10 [15]. Además, incluye etiquetas globales de puntuación Gleason y anotaciones de subtipos como el cribiforme, además de metadatos relevantes por paciente [7]. Esta base de datos es ampliamente utilizada en investigaciones recientes y está disponible para la comunidad científica bajo licencias abiertas, lo que garantiza su confiabilidad y legalidad en el uso académico. Su estructura también permite implementar validación cruzada estratificada por paciente, una estrategia crucial para asegurar la robustez del modelo y evitar el sobreajuste en entornos clínicos reales.

Adicionalmente, se garantiza que el conjunto de datos ofrece un número suficiente de muestras tanto en cantidad como en diversidad morfológica para entrenar modelos avanzados, lo cual es indispensable para una tarea de clasificación robusta y generalizable. Dado que las etiquetas son globales, el enfoque no requiere técnicas supervisadas tradicionales, pero se evaluará el rendimiento en distintas variantes arquitectónicas como atención, pooling jerárquico, VGPMIL-PR, todas compatibles con el volumen y la naturaleza de los datos disponibles.

Desde el punto de vista metodológico, el proyecto contempla no solo el desarrollo de modelos con alto rendimiento predictivo, sino también el uso de técnicas de explicabilidad como mapas de atención y arquitecturas con mecanismos tipo transformer, lo cual incrementa la confianza del profesional de la salud al

interpretar los resultados. En entornos clínicos, la interpretabilidad es tan crítica como la precisión, ya que permite integrar el modelo como herramienta complementaria en el proceso diagnóstico, y no como caja negra.

El impacto esperado de esta investigación es amplio y positivo. En primer lugar, podría reducir la carga de trabajo de los patólogos, acortar los tiempos de diagnóstico y aumentar la precisión en la identificación de subtipos histológicos de alto riesgo. En segundo lugar, el enfoque desarrollado tiene un alto grado de transferibilidad, al poder adaptarse a otros tipos de cáncer con estructuras histológicas similares, como el de mama o el colorrectal. Finalmente, al eliminar la dependencia de anotaciones exhaustivas, se propone una solución más accesible para centros clínicos con recursos limitados, contribuyendo a democratizar el acceso a tecnologías de diagnóstico basadas en inteligencia artificial.

Capítulo 8

Marco Teórico

8.1. Glosario

1. **Bag e Instancia:** En MIL, un *bag* es un conjunto de instancias (parches de WSI) etiquetadas colectivamente; cada instancia carece de etiqueta individual [3].
2. **Técnicas de agregación:** Métodos como *pooling* (max, mean, jerárquico) y mecanismos de atención asignan pesos a cada instancia para combinar sus características en una predicción de bolsa [4].
3. **VGPMIL-PR:** MIL probabilístico con Gaussian Processes y término Ising para modelar correlación espacial entre parches [6].
4. **Escala Gleason:** Sistema que gradúa la agresividad del cáncer de próstata según patrones glandulares [11].

8.2. Marco Conceptual

El estudio del cáncer de próstata desde la perspectiva de la ciencia de datos requiere comprender de manera conjunta los fundamentos clínicos de la enfermedad, los métodos tradicionales de evaluación histopatológica y las técnicas computacionales que permiten procesar y modelar grandes volúmenes de datos provenientes de imágenes digitales. En esta sección se revisan los conceptos centrales que sustentan la propuesta de investigación, organizados desde el dominio clínico hasta los marcos metodológicos de aprendizaje profundo que viabilizan su aplicación práctica.

8.2.1. Cáncer de próstata y escala de Gleason

El cáncer de próstata es una de las principales causas de morbilidad y mortalidad en la población masculina a nivel global, y su diagnóstico y pronóstico clínico dependen en gran medida de la evaluación histopatológica. El estándar de referencia para la clasificación de la agresividad tumoral es la escala de Gleason, la cual se basa en la observación de los patrones arquitecturales glandulares en las tinciones convencionales (H&E) de las biopsias. Esta escala asigna dos patrones (primario y secundario) con valores del 1 al 5, cuyo resultado combinado determina el puntaje final y la categoría de riesgo clínico [12] [1].

No obstante, la asignación de la puntuación de Gleason presenta una notable variabilidad interobservador. Esta subjetividad es particularmente crítica para patrones fronterizos (p. ej., Gleason 3 vs. 4) y para subtipos de peor pronóstico, como el patrón cribiforme del patrón 4, que se asocia con un mayor riesgo de recurrencia tumoral [11]. La variabilidad en la interpretación tiene implicaciones clínicas directas, ya que puede alterar la clasificación de riesgo, la indicación de tratamiento y los desenlaces a largo plazo. Este escenario clínico, que subraya la necesidad de mejorar la reproducibilidad diagnóstica, constituye el punto de partida ideal para vincular metodologías de ciencia de datos que operan sobre imágenes histológicas digitalizadas.

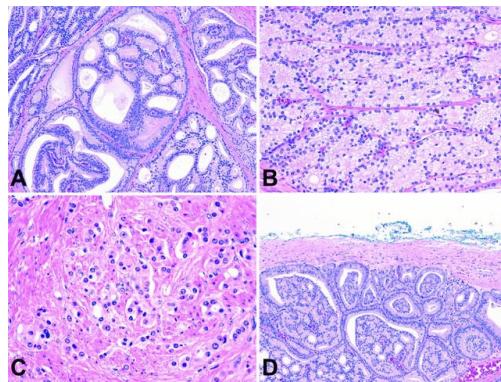


Figura 8.1: Ejemplos representativos de patrones cribiformes (Gleason 4) en tejido prostático [11].

8.2.2. Histopatología digital y Whole Slide Images (WSI)

La histopatología digital, impulsada por el escaneo de muestras histológicas, ha revolucionado el análisis de tejidos al generar Whole Slide Images (WSI). Estas imágenes gigapíxel, como se puede apreciar en la figura 8.1, conservan la arquitectura tisular completa y posibilitan el análisis computacional a gran escala, prometiendo asistir a los patólogos al reducir el error humano y estandarizar los criterios de evaluación [2] [5].

No obstante, las WSI plantean desafíos técnicos significativos debido a su tamaño extremo y la heterogeneidad visual inherente al tejido. Para hacer viable su procesamiento, estas imágenes se dividen en pequeños bloques o parches. Esta segmentación, sin embargo, introduce nuevos retos, como la variabilidad de tinción entre laboratorios, la presencia de artefactos de preparación y la complejidad morfológica intra e interpaciente [13].

Además, la ausencia de anotaciones detalladas a nivel de parche (instancia) hace que el Aprendizaje de Instancias Múltiples (MIL) sea un enfoque particularmente atractivo, ya que permite entrenar modelos con etiquetas globales a nivel de la imagen (slide-level), reduciendo la dependencia de anotaciones exhaustivas por parte de expertos. Por lo tanto, cualquier pipeline reproducible debe considerar cuidadosamente las decisiones de preprocesamiento, como la normalización de tinción y el filtrado por porcentaje de tejido, así como las estrategias de muestreo y representación que se adapten a estas complejidades [2] [5].

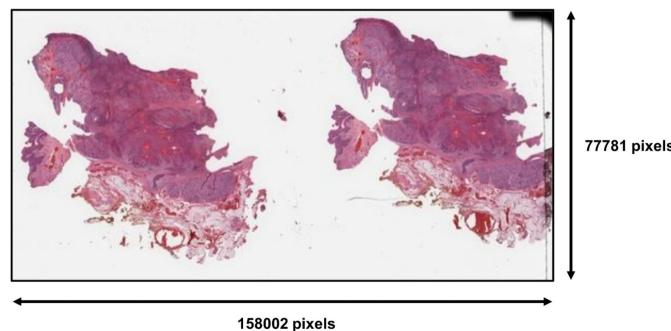


Figura 8.2: Ejemplo de imagen de super-resolución WSI [16]

8.2.3. Redes neuronales convolucionales como extractores de características

Las redes neuronales convolucionales (CNN) se han establecido como el paradigma dominante en la visión por computadora debido a su capacidad intrínseca para aprender jerarquías de características espaciales y patrones visuales complejos [17]. En el contexto de la histopatología digital, las CNN se emplean fundamentalmente como extractores de características a nivel de parche, lo cual es vital para procesar las imágenes de diapositivas completas (WSI) dada su gran escala. Cada parche individual es procesado por una CNN para obtener una representación vectorial (embedding) que encapsula la información morfológica, como las texturas y la arquitectura glandular, elementos clave para el diagnóstico de cáncer de próstata. Esta capacidad contribuye a reducir la subjetividad inherente a la interpretación visual humana [5] [18].

A continuación, se presenta la figura 8.3 que ilustra la arquitectura típica de una CNN.

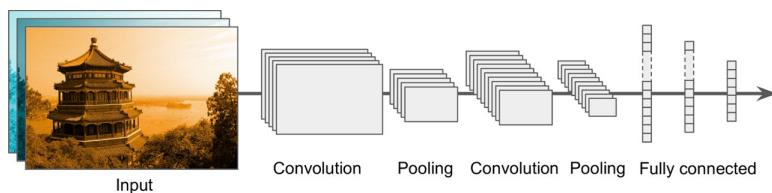


Figura 8.3: Arquitectura típica de una red neuronal convolucional [17]

Una estrategia comúnmente adoptada es el uso de transfer learning con arquitecturas pre-entrenadas en grandes bases de datos como ImageNet. Modelos como ResNet o EfficientNet, son adaptados para la tarea de clasificación histopatológica, lo cual acelera la convergencia del entrenamiento y mejora la generalización en conjuntos de datos de tamaño moderado [19] [20] [5]. Para estabilizar el proceso de aprendizaje y mitigar el sobreajuste, es crucial incorporar técnicas de regularización (como batch normalization y dropout) y estrategias de aumentación de datos diseñadas específicamente para el dominio de la histología.

En el marco del Aprendizaje de Instancias Múltiples (MIL), las características vectoriales extraídas por la CNN de cada parche son posteriormente agregadas por módulos especializados, como los mecanismos de atención o el pooling jerárquico. Esto permite generar una predicción a nivel de la lámina completa, eliminando la necesidad de un etiquetado exhaustivo de cada parche y consolidando el pipeline de análisis de WSI.

8.2.4. Aprendizaje de Instancias Múltiples (MIL)

El **Aprendizaje de Instancias Múltiples (MIL)** se presenta como una solución para escenarios en los que las etiquetas están disponibles únicamente a nivel de muestra completa, o “bag”. En este contexto, una WSI es interpretada como un *bag* de parches, y la tarea del modelo consiste en inferir la etiqueta global a partir de las características de las instancias locales. Si bien las primeras formulaciones se basaban en estrategias de agregación simples, como el *max pooling*, bajo la premisa de que la presencia de un solo parche positivo determinaba el diagnóstico, este enfoque ignoraba la heterogeneidad del tejido y las cruciales relaciones espaciales entre parches [3] [4].

La investigación reciente ha impulsado variantes más sofisticadas que abordan estas limitaciones. Entre ellas destacan los mecanismos de atención, que asignan pesos según la relevancia de cada parche; el pooling jerárquico, que captura información en múltiples escalas; y los modelos probabilísticos con regularización

espacial, diseñados para incorporar la correlación entre parches vecinos y estimar la incertidumbre de la predicción.

Estos avances no solo han mejorado la precisión diagnóstica del MIL, sino que también han potenciado su explicabilidad clínica, al permitir la identificación visual de las regiones histológicas más relevantes para el diagnóstico final.

8.2.5. Métricas de evaluación e interpretabilidad

La validación rigurosa de modelos en histopatología digital va más allá del simple desempeño numérico, requiriendo métricas que reflejen tanto la utilidad técnica como la relevancia clínica del sistema [17]. En este contexto, la evaluación se basa en un conjunto de indicadores clave, cada uno de los cuales ofrece una perspectiva diferente sobre el comportamiento del modelo:

- **AUC (Área bajo la curva ROC):** evalúa la capacidad de discriminación del modelo para distinguir entre las clases positivas y negativas a través de todos los posibles umbrales de clasificación. Un valor de AUC cercano a 1 indica una excelente capacidad para clasificar correctamente los casos.
- **Precisión:** mide la proporción de diagnósticos positivos correctos sobre el total de casos clasificados como positivos. En un escenario clínico, una alta precisión es crucial, ya que minimiza los falsos positivos que podrían llevar a tratamientos innecesarios.
- **Recall o sensibilidad:** cuantifica la proporción de casos positivos que el modelo identificó correctamente. Una alta sensibilidad es vital para evitar los falsos negativos, que representan el riesgo de no detectar una enfermedad en su etapa inicial.
- **F1-score:** representa la media armónica entre precisión y recall. Esta métrica es especialmente útil en conjuntos de datos desbalanceados, ya que proporciona un solo valor que resume el equilibrio entre evitar falsos positivos y falsos negativos.
- **Coeficiente Kappa de Cohen:** cuantifica el grado de acuerdo entre las predicciones del modelo y las etiquetas reales, corrigiendo el acuerdo que podría ocurrir por simple azar. A diferencia de la precisión simple, el coeficiente Kappa es particularmente relevante para comparar el desempeño del modelo con el grado de acuerdo interobservador entre especialistas humanos.

Más allá de las métricas numéricas, la interpretabilidad es esencial para la adopción clínica. Herramientas como mapas de activación (ej. Grad-CAM) y mecanismos de atención en MIL permiten visualizar las regiones que sustentan la decisión del modelo. Complementariamente, la estimación de incertidumbre ayuda a distinguir entre predicciones confiables y casos que requieren revisión experta [17].

Capítulo 9

Marco de Antecedentes

La investigación en aprendizaje débilmente supervisado sobre Whole Slide Images (WSI) ha avanzado rápidamente en los últimos años, motivada por la necesidad de reducir la carga de anotación experta y por el potencial clínico de modelos reproducibles y explicables. Varios trabajos han propuesto arquitecturas y estrategias que permiten agregar evidencia local (parches) en predicciones a nivel de muestra (bag), y han explorado alternativas para incorporar contexto espacial, medidas de incertidumbre y visualizaciones de interpretabilidad.

Para contextualizar la presente investigación, a continuación, se revisan los antecedentes más relevantes que abordan estas problemáticas. Dichos estudios justifican las decisiones metodológicas de este proyecto al ilustrar las principales tendencias y desafíos en el campo, sirviendo como base teórica para el diseño de la solución propuesta:

9.0.1. CAMIL: channel attention-based multiple instance learning for whole slide image classification

En este trabajo, los autores presentan una arquitectura denominada FA-MILNet (Feature Aggregation Multiple Instance Learning Network), diseñada para abordar los desafíos del aprendizaje débilmente supervisado en la clasificación de imágenes histopatológicas de diapositivas completas (WSI). La propuesta surge como respuesta a las limitaciones de los métodos tradicionales de aprendizaje de instancias múltiples (MIL), los cuales suelen perder información relevante al aplicar estrategias simples de agregación como el max-pooling, además de ignorar las relaciones contextuales entre parches dentro de la imagen. FA-MILNet introduce un enfoque de fusión jerárquica que combina características locales y globales mediante un módulo de atención que identifica regiones clave dentro de cada WSI. El modelo se estructura en tres bloques: extracción de características por CNN, mecanismo de atención y una rama dual de agregación, lo que permite mejorar tanto la precisión como la interpretabilidad [21].

Evaluado en conjuntos de datos de cáncer de colon, mama y ganglios linfáticos (Camelyon16, BACH, CRC), el modelo superó a métodos establecidos como ABMIL, CLAM y DSMIL en métricas como AUC y F1-score. Aunque el enfoque no se aplicó directamente a cáncer de próstata, su diseño modular y su capacidad para operar sin anotaciones a nivel de parche, lo convierten en una referencia altamente pertinente para proyectos que buscan aplicar MIL en entornos histopatológicos complejos y clínicamente realistas, demostrando el valor de los mecanismos de atención para ponderar la relevancia de los parches, una estrategia central en nuestra propuesta, y subraya la importancia de una arquitectura de agregación sofisticada que no se limite a operaciones simples. [21].

9.0.2. A Probabilistic MIL Model with Spatial Regularization for Weakly Supervised Histopathology Image Analysis

Este artículo se centra en la propuesta de un modelo MIL basado en procesos gaussianos, que incorpora una regularización espacial inspirada en el modelo de Ising. Su objetivo fundamental es capturar la dependencia entre parches vecinos dentro de una WSI, partiendo de la premisa de que la cercanía entre regiones histológicas implica una alta correlación diagnóstica. Un aspecto crucial de este enfoque es su capacidad para estimar distribuciones de probabilidad sobre las etiquetas de los parches, lo que a su vez permite cuantificar la incertidumbre en las predicciones, una cualidad indispensable en entornos clínicos donde la fiabilidad del modelo impacta directamente en las decisiones médicas [6].

A diferencia de otros modelos determinísticos, la naturaleza probabilística de esta propuesta facilita una interpretación más robusta de las regiones relevantes, logrando identificar zonas tumorales con una mayor precisión incluso bajo supervisión débil. La principal contribución de este trabajo radica en la demostración de que la combinación de incertidumbre y regularización espacial puede mejorar de manera significativa la interpretabilidad y la fiabilidad de las predicciones a nivel de parche. Este concepto es fundamental para nuestro proyecto, ya que la incorporación de información espacial y la estimación de la incertidumbre son pilares esenciales para construir un modelo no solo preciso, sino también robusto y clínicamente útil [6].

9.0.3. Graph-based Multiple Instance Learning for Whole Slide Image Classification

Este estudio introduce una variante de MIL que modela explícitamente las relaciones espaciales entre parches de una WSI mediante un grafo no dirigido. Cada parche es representado como un nodo, y las conexiones entre nodos reflejan la vecindad espacial entre regiones del tejido. La arquitectura incorpora una red neuronal de grafos (GNN) que permite propagar información contextual entre los nodos, fortaleciendo la capacidad del modelo para interpretar patrones arquitectónicos globales. Esta aproximación ha demostrado ser particularmente efectiva para subtipos tumorales que no se definen por características locales aisladas, sino por distribuciones estructurales complejas [22].

Los resultados presentados indican mejoras importantes en tareas de clasificación multicategoría frente a modelos MIL convencionales. Este trabajo es un antecedente crucial porque valida la hipótesis de que las relaciones espaciales son informativas y deben ser incorporadas explícitamente en el modelo. A diferencia de los enfoques de atención que ponderan la relevancia de los parches de forma individual, la aproximación basada en grafos permite una propagación de información contextual a través de toda la WSI, lo cual podría ser una estrategia complementaria o alternativa para nuestra investigación [22].

9.0.4. Deep Recurrent Attention MIL for Histopathological Image Analysis

El enfoque planteado en este artículo se distingue por combinar mecanismos de atención con redes neuronales recurrentes (RNN) dentro del marco de MIL. En lugar de tratar los parches como elementos independientes, los autores introducen una secuencia ordenada basada en la posición espacial de cada uno, la cual es procesada por una RNN con atención. Este diseño permite capturar no solo las características locales, sino también las relaciones de largo alcance y las dependencias contextuales más profundas a lo largo de toda la WSI [23].

Gracias a esta arquitectura secuencial, el modelo demuestra un mejor rendimiento en la identificación de estructuras glandulares complejas, un desafío particular en escenarios con gran variabilidad morfológica

entre pacientes. Además, el mecanismo de atención adaptativa complementa este diseño, facilitando interpretaciones más precisas del modelo al resaltar de forma dinámica las regiones más relevantes. La principal lección de este trabajo es que la secuencia y el orden de los parches, y no solo su presencia, pueden aportar información crucial. A diferencia de los enfoques basados en grafos que modelan la vecindad de forma explícita, este método explora una representación secuencial, lo que invita a considerar cómo la estructura de los datos de entrada puede ser modificada para extraer de manera más efectiva la información contextual [23].

9.0.5. Hierarchical Pooling in MIL for Histopathological Subtype Classification

Este trabajo propone un esquema de agregación jerárquica dentro del paradigma MIL, en el cual los parches primero se agrupan en regiones intermedias según su proximidad o similitud morfológica. Posteriormente, se aplica un segundo nivel de pooling que combina la información de estas regiones para obtener la predicción final del bag. Esta estructura multinivel permite capturar patrones de organización tisular más amplios, que son clave en el diagnóstico de ciertos subtipos de cáncer. La metodología incorpora tanto técnicas de agrupamiento automático como atención regional, y ha mostrado un desempeño superior en conjuntos de datos con alta heterogeneidad intratumoral, al mantener el balance entre granularidad local y contexto estructural [24].

Este antecedente es relevante porque, al igual que FA-MILNet, enfatiza la importancia de una agregación multinivel. Sin embargo, se distingue por su énfasis en la creación de “grupos” lógicos de parches antes de la agregación final, lo que es una estrategia diferente a la simple ponderación por atención. Por otro lado, la idea de agrupar parches basándose en la morfología o la proximidad podría ser una técnica poderosa a considerar para proyectos que buscan capturar la complejidad de la heterogeneidad tumoral [24].

Capítulo 10

Metodología

La metodología adoptada para el desarrollo del presente proyecto es el modelo CDIO (Concebir, Diseñar, Implementar y Operar), un marco iterativo y orientado a ingeniería que facilita la trazabilidad entre objetivos, tareas y entregables, así como la incorporación de ciclos de retroalimentación para la mejora continua [25]. A continuación, se describen las tareas que se realizarán en cada una de las etapas anteriores, para poder cumplir con los objetivos específicos planteados en un principio.

Objetivo 1: Implementar técnicas de procesamiento de imágenes de histopatología prostática para resaltar características diagnósticas y optimizar la calidad de los parches de entrada para el modelo.

- 1.1 **Concebir:** Revisar sistemáticamente la literatura sobre preprocesamiento de WSI y seleccionar técnicas adecuadas para el dominio prostático.
- 1.2 **Diseñar:** Definir el pipeline de preprocesamiento y extracción de parches y la estructura de metadatos que acompañará a cada parche.
- 1.3 **Implementar:** Desarrollar código reproducible para normalización de color, eliminación de fondo, filtrado de parches, aumentos específicos para histología y, opcionalmente, métodos de mejora de resolución o contraste.
- 1.4 **Operar:** Medir cuantitativamente la calidad de los parches mediante métricas de cobertura de tejido, estadísticas de color.

Objetivo 2: Implementar técnicas de visión por computadora basadas en aprendizaje de múltiples instancias (MIL) con Deep Learning para agregar las representaciones de los parches y producir etiquetas a nivel de Whole Slide Image.

- 2.1 **Concebir:** Seleccionar las variantes de MIL a implementar y definir criterios de diseño y comparación (capacidad de explicación, costo computacional, complejidad de entrenamiento).
- 2.2 **Diseñar:** Especificar la arquitectura final del pipeline de aprendizaje: extractor de características (CNN pre-entrenada o entrenamiento desde cero), módulo de agregación, estrategia de optimización y esquema de checkpointing.
- 2.3 **Implementar:** Programar las arquitecturas seleccionadas en PyTorch y/o TensorFlow, integrar OpenSlide para lectura eficiente de WSI, preparar scripts de entrenamiento, validación y prueba.
- 2.4 **Operar:** Optimizar el pipeline mediante la búsqueda y ajuste de hiperparámetros con validación interna y criterios de early stopping, complementado con pruebas de escalabilidad y rendimiento en los entornos de cómputo disponibles.

Objetivo 3: Evaluar el rendimiento de los modelos desarrollados mediante métricas estándar (AUC, precisión, recall, F1-score y coeficiente Kappa), empleando validación cruzada.

- 3.1 **Concebir:** Definir el protocolo experimental (validación cruzada estratificada por paciente, número de folds, semillas reproducibles, criterios de selección de conjunto de prueba) y seleccionar las métricas primarias y secundarias a reportar.
- 3.2 **Diseñar:** Establecer el pipeline de evaluación automática que incluya generación de curvas ROC, matrices de confusión, cálculo del coeficiente Kappa por agrupación Gleason y/o procedimientos estadísticos para comparar modelos.
- 3.3 **Implementar:** Ejecutar los experimentos planificados para cada variante de modelo, recolectar métricas y metadatos de ejecución y almacenar resultados en formato reproducible.
- 3.4 **Operar:** Realizar análisis cuantitativo y/o cualitativo de resultados, incluyendo comparaciones estadísticas entre métodos, visualizaciones y revisión de errores relevantes.

Capítulo 11

Recursos

11.1. Recursos

Para la ejecución del proyecto se identifican y detallan a continuación los recursos humanos, equipos, software, datos y servicios necesarios. Se indica la titularidad o el origen de cada recurso (Javeriana = aportado por la universidad; Recursos propios = aportado por los integrantes o laboratorio) y la disponibilidad o mecanismo para su acceso. No se incluyen tablas de presupuesto monetario en cumplimiento de la observación del evaluador; la estimación económica podrá entregarse por separado si se requiere.

11.1.1. Participantes

- **María Valentina Belalcázar Perdomo:** Ingeniera Ambiental con especialización en Seguridad y Salud en el Trabajo, analítica de datos y sólida trayectoria en auditoría interna, implementación y mejora continua de Sistemas Integrados de Gestión (SIG) bajo estándares de calidad, medio ambiente, seguridad y salud ocupacional.
- **Juan José Restrepo Rosero:** Ingeniero Electrónico, con experiencia en desarrollo de software DevOps. Competencias técnicas en Python, PyTorch y despliegue de pipelines locales.
- **Dr. Julián Gil González:** Director del proyecto, profesor catedrático de la Pontificia Universidad Javeriana - Cali. Doctor en Ingenierías y Magíster en Ingeniería Eléctrica, con especialización en Reconocimiento de Patrones y Aprendizaje Automático por la Universidad Tecnológica de Pereira. Cuenta con más de diez años de experiencia en el desarrollo de modelos de Aprendizaje Automático. Posee una sólida trayectoria en visión por computadora y procesamiento del lenguaje natural. Ha diseñado e implementado modelos de AA utilizando TensorFlow y PyTorch, y ha participado en proyectos colaborativos relacionados con aplicaciones de inteligencia artificial en ámbitos como la bioingeniería y los sistemas de transporte masivo.

11.1.2. Descripción general de recursos materiales y su disponibilidad

A continuación, se listan los recursos materiales y su origen. Se especifica la forma de acceso y una nota sobre disponibilidad.

▪ Equipos de cómputo:

- 2 estaciones de trabajo Asus TUF equipadas para entrenamiento local (GPU compatible). *Origen: Recursos propios. Disponibilidad: confirmada por los integrantes para desarrollo y pruebas iniciales.*
- Infraestructura de cómputo de la Facultad / Laboratorio (acceso a servidores y/o GPUs según políticas de la universidad). *Origen: Javeriana. Disponibilidad sujeta a coordinación con la Facultad.*

- Opción de instancias en la nube (AWS/GCP/Azure) para entrenamientos a gran escala — a cotizar y utilizar sólo si las pruebas locales indican necesidad. *Origen: externo / por contratar si aplica.*

■ **Software y herramientas:**

- Lenguajes y bibliotecas: Python, PyTorch, TensorFlow, scikit-learn, OpenSlide, tiffile, OpenCV, pandas, matplotlib, etc. *Origen: Software libre / provisto por la universidad.*
- Gestión y reproducibilidad: Git (repositorio en GitHub), environment.yml / requirements.txt, scripts de ejecución y notebooks reproducibles. *Origen: Recursos propios.*
- Servicios para seguimiento y registro (opcional): Weights & Biases, TensorBoard. *Origen: externos; uso según disponibilidad y presupuesto.*

■ **Datos:**

- **SICAPv2** — Prostate Whole Slide Images with Gleason grades (dataset público). Contiene WSI y metadatos (155 WSI, 95 pacientes, ~18,783 parches etiquetados). *Origen: público; licencia académica/uso para investigación.* Se adjunta referencia y enlace en la bibliografía.
- (Si aplica) Datos clínicos adicionales o colaboraciones con instituciones: se documentarán y se presentarán autorizaciones/consentimientos según la normativa ética vigente antes de su uso.

■ **Otros servicios y apoyo:**

- Acceso a bibliotecas científicas y bases de datos institucionales (PUJC). *Origen: Javeriana.*
- Soporte de TI para instalación de software y acceso a recursos de la Facultad (según proceda). *Origen: Javeriana.*

11.1.3. Aseguramiento de acceso y validación de disponibilidad

- **Acceso a datos:** SICAPv2 es público y se dispone del enlace oficial para descarga; se documentará el procedimiento de obtención y la fecha de descarga en los metadatos del repositorio del proyecto.
- **Acceso a cómputo:** Las estaciones de trabajo propias permitirán las pruebas iniciales y el desarrollo del pipeline. Para experimentos que requieran mayor capacidad se solicitará acceso a recursos de la Facultad y/o uso controlado de instancias en la nube (previa justificación técnica y, en caso de requerir financiación, cotización).
- **Documentación y reproducibilidad:** Se mantendrá un repositorio Git con control de versiones, scripts automatizados para generación de parches, preprocessamiento, entrenamiento y evaluación, además de ficheros de entorno (environment.yml / requirements.txt) y semillas de aleatoriedad para garantizar trazabilidad.

11.1.4. Entregables relacionados con recursos

- Pipeline de preprocessamiento y extracción de parches (documentado).
- Repositorio de código con implementaciones reproducibles y checkpoints de modelos (si aplicable).
- Registro de metadatos: splits por paciente, rutas de parches y versiones de datasets.
- Informe de disponibilidad y validación de recursos (incluye confirmación de acceso a la infraestructura universitaria cuando sea necesario).

Capítulo 12

Cronograma

12.0.1. Cronograma

El tiempo total del proyecto se llevará a cabo en 24 semanas las cuales se distribuyen de la siguiente manera:

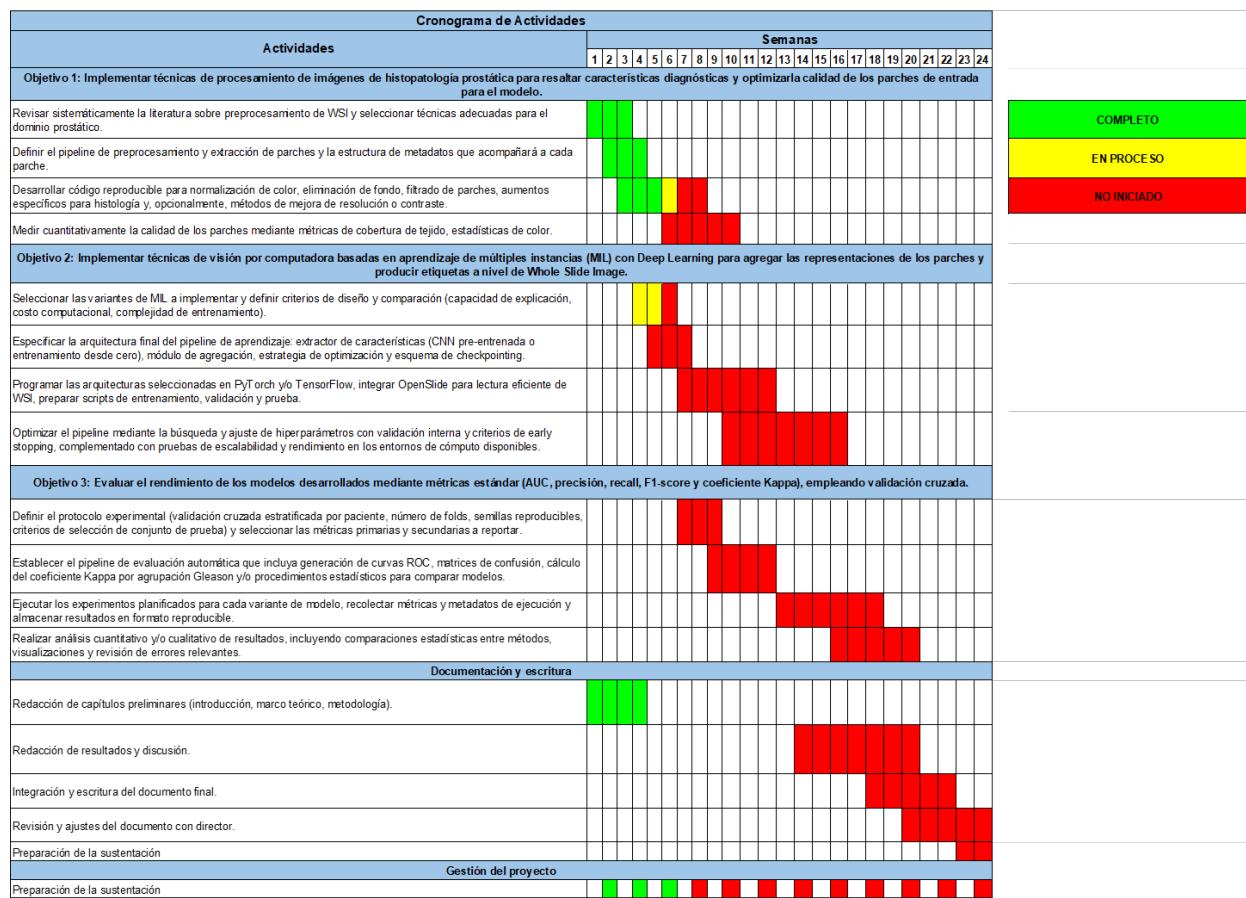


Figura 12.1: Cronograma del proyecto (Elaboración propia)

Bibliografía

- [1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [2] G. Campanella, M. G. Hanna, L. Geneslaw, and et al., “Clinical-grade computational pathology using weakly supervised deep learning on whole slide images,” *Nature Medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.
- [3] M.- Carboneau, V. Cheplygina, E. Granger, and G. Gagnon, “Multiple instance learning: A survey of problem characteristics and applications,” *Pattern Recognition*, vol. 77, pp. 329–353, 2018.
- [4] M. Ilse, J. M. Tomczak, and M. Welling, “Attention-based deep multiple instance learning,” *arXiv preprint arXiv:1802.04712*, Jun 2018, accessed: 2025-09-1.
- [5] M. Lu *et al.*, “Data-e"cient and weakly supervised computational pathology on whole-slide images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Mar 2021, pp. 8359–8369.
- [6] P. Morales-Álvarez, A. Schmidt, J. M. Hernández-Lobato, and R. Molina, “Introducing instance label correlation in multiple instance learning. application to cancer detection on histopathological images,” *Pattern Recognition*, vol. 146, p. 110057, 2024.
- [7] J. Silva-Rodríguez, “Sicapv2 - prostate whole slide images with gleason grades annotations,” Oct 2020, available: <https://data.mendeley.com/datasets/9xxm58dvs3/2>.
- [8] F. Ciompi and et al., “Standardized evaluation of algorithms for computer-aided diagnosis of melanoma and nonmelanoma skin cancer based on the ham10000 dataset,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1292–1299, 2017.
- [9] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, “Global cancer statistics 2020:globocan estimates of incidence and mortality,” 2021, summary available from: <https://pubmed.ncbi.nlm.nih.gov/33538338/>. Accessed: 2025-09-23.
- [10] M. MinSalud, “Observatorio nacional de cáncer >indicadores >incidencia cáncer de próstata en colombia,” 2025, available from: <https://www.sispro.gov.co/observatorios/oncancer/indicadores/Paginas/C>
- [11] J. Silva-Rodríguez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, “Going deeper through the gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection,” *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105637, 2020, available: <https://www.sciencedirect.com/science/article/pii/S016926072031470X>
- [12] B. Delahunt, R. J. Miller, J. R. Strigley, A. J. Evans, and H. Samaratunga, “Gleason grading: Past, present and future,” Sep 2012, available: <https://pubmed.ncbi.nlm.nih.gov/22212079/>.
- [13] D. Tellez, “Whole-slide mitosis detection in the breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks,” 2018, available: <https://pubmed.ncbi.nlm.nih.gov/29994086/>.

- [14] T. A. Ozkan, A. T. Eruyar, O. O. Cebeci, O. Memik, L. Ozcan, and I. Kuskonmaz, "Interobserver variability in Gleason histological grading of prostate cancer," *Scandinavian Journal of Urology*, vol. 50, no. 6, pp. 420–424, 2016.
- [15] X. Zhao, Y. Zhang, X. Liang *et al.*, "Prostate gland segmentation study using swin transformer," *Proceedings of SPIE—the International Society for Optical Engineering*, vol. 12627, p. 126270N, 2023, dataset statistics for SICAPv2: 155 WSI, 95 patients, 18,783 labeled patches, available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11640496/>.
- [16] J. Yao, "Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks," *Medical Image Analysis*, Oct 2020, accessed 11-09-2025.
- [17] A. Géron, *Hands-on machine learning with scikit-learn, Keras, and tensorflow*, 3rd ed. O'Reilly, 2022.
- [18] H. Hezi, M. Gelber, A. Balabanov, Y. E. Maruvka, and M. Freiman, "Cimil-crc: A clinically-informed multiple instance learning framework for patient-level colorectal cancer molecular subtypes classification from he stained images," *Computer Methods and Programs in Biomedicine*, vol. 259, p. 108513, 2025, available: <https://www.sciencedirect.com/science/article/pii/S0169260724005066>.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [20] M. Tan and Q. Le, "E"cientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.
- [21] J. Mao, J. Xu, X. Tang, Y. Liu, H. Zhao, G. Tian, and J. Yang, "Camil: channel attention-based multiple instance learning for whole slide image classification," *Bioinformatics*, vol. 41, no. 2, Feb. 2025, publisher Copyright: © The Author(s) 2025. Published by Oxford University Press.
- [22] O. Fourkioti, M. De Vries, C. Jin, D. C. Alexander, and C. Bakal, "Context-aware multiple instance learning for cancer detection and subtyping in whole slide images," Oct 2024, available: <https://arxiv.org/abs/2305.05314>. Accessed: 2025-08-11.
- [23] L. Qu, D. Yang, D. Huang, Q. Guo, R. Luo, S. Zhang, and X. Wang, "Pathology-knowledge enhanced multi-instance prompt learning for few-shot whole slide image classification," *Pathology-knowledge enhanced multi-instance prompt learning for few-shot whole slide image classification*, Jul 2024, available: <https://arxiv.org/abs/2407.10814>. Accessed: 2025-08-11.
- [24] M. Waqas, S. U. Ahmed, M. A. Tahir, J. Wu, and R. Qureshi, "Exploring multiple instance learning (mil): A brief survey," *Expert Systems with Applications*, vol. 250, p. 123893, 2024, available <https://www.sciencedirect.com/science/article/pii/S0957417424007590>. Accessed: 2025-08-11.
- [25] E. Amaya, "Metodologías de trabajo," *Metodologías de Trabajo*, 2015, available: <https://isfcolombia.uniandes.edu.co/index.php/eventos/metodologias-de-trabajo>.