



Proyecto Final de Carrera de Bioingeniería

**Desarrollo de una herramienta de patología digital para la
clasificación de carcinomas endometriales proficientes y
deficientes del sistema de reparación del ADN, mediante un
algoritmo de inteligencia artificial. Su utilidad clínica y
aplicaciones futuras**

Julieta Chirkes (60070)

Ana Gorodisch (60274)

Tutora: Dra. María Teresa Pombo

Co-tutor: Lic. Adrián Pérez

Fecha de entrega: 15/11/2023

Agradecimientos

El presente trabajo no habría sido posible sin el apoyo incondicional de nuestros seres queridos. Muchas gracias a nuestra familia y amigos por acompañarnos durante este proceso; por momentos arduo y frustrante, y por otros sumamente gratificante. Su constante aliento fue crucial. Una mención especial a Andrea, Martina, Nacho y Olivia, por haber sido excelentes compañeros estos cinco años de carrera. Gracias a la Licenciada Belén Villegas, por su colaboración desinteresada en la digitalización de las imágenes; Gracias infinitas al Licenciado Adrián Pérez, por su dedicación e inspiración en el proyecto. Nada de esto hubiese sido posible sin él. Finalmente, muchísimas gracias a la Doctora Teresa Pombo, por ser la impulsora de esta tesis. Día y noche, siempre estuvo disponible en pos de que este trabajo agregue valor al campo de la medicina. Ha sido un gran honor trabajar a su lado.

Índice general

Índice de figuras	v
Índice de tablas	viii
1. Resumen	1
2. Introducción	3
2.1. Objetivos del trabajo	4
3. Marco teórico	6
3.1. Cáncer de endometrio	6
3.1.1. Incidencia y mortalidad	8
3.1.2. Factores de riesgo	9
3.1.3. Carcinoma de endometrio hereditario	10
3.1.4. Clasificación del cáncer de endometrio	10
3.1.5. Diagnóstico	12
3.2. Inestabilidad microsatelital y cáncer de endometrio	13
3.2.1. Microsatélites	13
3.2.2. Mecanismo de reparación del ADN	14
3.2.3. Inestabilidad microsatelital y sus variantes somática y germinal	15
3.2.4. Diagnóstico de inestabilidad microsatelital	16
3.2.5. Inestabilidad microsatelital y su mejor pronóstico	19
3.3. Tratamiento	21
3.3.1. Terapia dirigida	22
3.4. Inteligencia artificial y aprendizaje profundo	23
3.4.1. Redes neuronales	23
3.4.2. Proceso de aprendizaje	27
3.4.3. Distintos tipos de entrenamiento	29
3.4.4. Aprendizaje por transferencia	31
3.5. Evaluación y métricas	33
3.6. Patología digital	35

3.7. Arquitectura CLAM	37
4. Desarrollo	39
4.1. Elección de la arquitectura CLAM como base para el modelo	41
4.2. Obtención de muestras	42
4.2.1. Conjunto TCGA-CPTAC	42
4.2.2. Imágenes de archivo de un laboratorio de patología	46
4.3. Hardware y software	47
4.4. Etapas del flujo de trabajo	48
4.4.1. Segmentación del tejido	48
4.4.2. Obtención de parches	50
4.4.3. Extracción de características	51
4.4.4. Proceso de aprendizaje	52
4.5. <i>Heatmaps</i> como herramienta visual	57
4.5.1. Análisis de los parches más relevantes para cada clase	58
4.5.2. Localización del tumor	58
4.6. Exploración de otras líneas de trabajo	61
4.7. Desarrollo de un prototipo de interfaz gráfica	63
5. Resultados	64
5.1. Selección del umbral óptimo para la tarea de clasificación	66
5.2. Evaluación del modelo con el conjunto independiente	68
5.3. Evaluación de la atención	69
5.4. Interpretación biológica de los parches distintivos para cada clase	71
5.5. Interfaz gráfica	72
6. Discusión	74
6.1. Comparación entre resultados dependientes e independientes	76
6.2. Comparación con el estado del arte	79
6.3. Futuras direcciones	81
7. Conclusión	83
8. Abreviaturas	85

9.	Anexo	87
9.1.	Clasificación histológica del cáncer de cuerpo de útero según la OMS	87
9.2.	Parámetros para la segmentación	88
9.3.	Cálculo del puntaje de atención para cada parche	88
9.4.	Optimizador Adam	89
9.5.	Matrices de confusión y curvas PR antes de la umbralización	90
10.	Referencias	93

Índice de figuras

3.1. Anatomía del útero [4]	6
3.2. Histología del útero. Endometrio, miometrio y perimetrio [7]	7
3.3. Tasas de incidencia de cáncer de cuerpo de útero estandarizadas por edad estimadas a nivel mundial en el año 2020 para todas las mujeres de todas las edades [9]	8
3.4. Tasas de incidencia y de mortalidad de los 10 cánceres más frecuentes estandarizadas por edad estimadas a nivel mundial en 2020 para todas las mujeres de todas las edades. Comparación entre Argentina y Estados Unidos [9]	9
3.5. Clasificación molecular del cáncer de endometrio [14]	12
3.6. Microsatélite dinucleótido	13
3.7. Replicación normal de un microsatélite (A), deslizamiento hacia adelante (B) y deslizamiento hacia atrás (C) [17]	14
3.8. Mecanismo MMR y las proteínas involucradas [20]	15
3.9. Flujo de trabajo del proceso de inmunohistoquímica (ejemplificado con inmunohistoquímica indirecta) [26]	17
3.10. Proteína MSH2 presente (izquierda) y ausente (derecha) en núcleos tumorales mediante IHQ [25]	18
3.11. Tumor "frío" y tumor "caliente" [31]	20
3.12. Esquema simplificado de una red neuronal	24
3.13. Esquemmatización de una red neuronal convolucional [48]	27
3.14. Tareas de aprendizaje supervisado y no supervisado	29
3.15. Esquemmatización de aprendizaje de múltiples instancias [58]	30
3.16. Esquemmatización de un ViT [63]	33
3.17. Matriz de confusión	33
3.18. Curva ROC y curva PR [66]	35
3.19. WSI de cáncer de endometrio	36
4.1. Flujo de trabajo de la herramienta completa	40
4.2. Ejemplos de imágenes con marcador, bálsamo y grandes pliegues	44

4.3. Comparación entre una buena (izquierda) y una mala (derecha) digitalización a una magnificación de 10x	47
4.4. Ejemplo de una resección	49
4.5. Ejemplo de una biopsia	49
4.6. Segmentación del tejido. Obtención de la máscara y aplicación de la misma.	50
4.7. Obtención de parches a partir de la región de tejido de la diapositiva	50
4.8. Comparación entre los parches extraídos a 10 y 20x	51
4.9. Proceso de aprendizaje del clasificador a nivel de imagen a través de la red de atención	54
4.10. Funcionamiento de la capa de agrupamiento ejemplificada para una imagen MSI-H	56
4.11. Comparación entre marcaciones de tumor realizadas por una patóloga y <i>heatmaps</i> para la localización del tejido tumoral	60
4.12. Comparación entre marcaciones de tumor realizadas por una patóloga y <i>heatmaps</i> para la localización del tejido tumoral para el caso de una infiltración linfocitaria (marcación amarilla)	61
4.13. Flujo de trabajo para la utilización de ViTs como modelo de clasificación	62
5.1. Curva PR del modelo_balanceado_10x	66
5.2. Posibles umbrales de decisión para el modelo_balanceado_10x	66
5.3. Matrices de confusión de los 4 modelos luego de la umbralización	67
5.4. Matriz de confusión del modelo_balanceado_10x evaluado con imágenes del conjunto independiente	68
5.5. Curva PR del modelo_balanceado_10x para el conjunto independiente	69
5.6. Posibles umbrales de decisión para el modelo_balanceado_10x con el conjunto independiente	69
5.7. Comparación entre las regiones en las que el algoritmo pone atención y las regiones tumorales	70
5.8. En el modelo digitalizado, parches con elevado puntaje de atención para cada clase	71
5.9. Prototipo de Interfaz Gráfica	73
6.1. Imágen del conjunto TCGA-CPTAC	77
6.2. Imágen del conjunto independiente	77

9.1. Clasificación histológica de la OMS para el cáncer de endometrio [85]	87
9.2. Matrices de confusión para los cuatro modelos antes de la umbralización óptima	90
9.3. Curva PR y posibles umbrales para el modelo_desbalanceado_20x . . .	91
9.4. Curva PR y posibles umbrales para el modelo_desbalanceado_10x . . .	91
9.5. Curva PR y posibles umbrales para el modelo_balanceado_20x	92

Índice de tablas

3.1. Desempeño de arquitectura CLAM en distintas tareas	38
4.1. Distintos modelos de entrenamiento	39
4.2. Subconjuntos de entrenamiento, validación y evaluación para los modelos entrenados con TCGA-CPTAC (balanceado y desbalanceado) .	46
5.1. Métricas para evaluar el desempeño de los distintos modelos	64
5.2. Métricas de cada uno de los modelos luego de la umbralización	67
5.3. Métricas del modelo_balanceado_10x evaluado en el conjunto independiente	68
6.1. Métricas estadísticas de los modelos <i>Panoptes</i> , <i>GAM-VGG11</i> , <i>im4MEC</i> y Modelo_balanceado_10x	80
9.1. Parámetros para la segmentación del tejido para cada imagen	88

Resumen

La inteligencia artificial (IA) busca replicar la inteligencia humana en procesos computacionales, creando un campo dentro de las ciencias de la computación que tiene como objetivo la resolución de problemas. Se puede implementar de manera eficiente mediante el desarrollo de diversos algoritmos adaptados a las necesidades del usuario. La patología digital, una excelente herramienta de trabajo en ciencias de la salud, es un entorno dinámico basado en imágenes que permite la adquisición, gestión e interpretación de información patológica generada a partir de un portaobjetos digitalizado [1]. Estas imágenes digitales pueden luego ser almacenadas, compartidas y analizadas digitalmente por los diferentes patólogos y otros profesionales afines, expertos en este área de trabajo.

El objetivo de este trabajo es utilizar IA para la predicción de inestabilidad microsatelital en carcinomas endometriales, desarrollando un algoritmo que nos permita separar, de este universo de pacientes, al subgrupo de tumores con sistema de reparación del ADN conservado (Proficiente, subrogante de tumores estables o MSS) de aquéllos con el sistema de reparación del ADN deficiente o averiado (Deficiente, subrogante de tumores inestables o MSI). La elección de tumores endometriales para nuestro proyecto radica en la importancia de reconocer el subgrupo MSI, por la respuesta a novedosos tratamientos con inmunoterapia entre otras consideraciones y el aporte invaluable en la simplificación de procesos por parte de la patología digital que describiremos a lo largo de nuestro trabajo.

El algoritmo desarrollado se basó en la arquitectura CLAM (*Clustering Constrained Attention Multiple Instance Learning*) que es un modelo de clasificación de imágenes patológicas, basado en la atención. Este modelo utiliza aprendizaje profundo en un marco semi supervisado en donde se entrena con WSI (*Whole Slide Images*) que tienen una única etiqueta (MSI o MSS) para toda la diapositiva. El modelo logra enfocarse en las regiones relevantes de la imagen para luego poder predecir la inestabilidad o estabilidad microsatelital en imágenes desconocidas para el modelo. En primer lugar, se segmentan las regiones de tejido dentro de la dia-

positiva y se divide la misma en parches más pequeños de 256 x 256 píxeles. Para reducir el costo computacional del modelo, se utiliza una red neuronal convolucional (CNN) para extraer características relevantes de cada uno de los parches. Son estas características las que se utilizan como entrada para el modelo de clasificación. Se entrenó al modelo con dichas características correspondientes a 397 WSI, 169 MSI y 228 MSS obtenidas de los proyectos públicos TCGA-UCEC y CPTAC-UCEC. Cuando se evaluó el modelo en 38 imágenes de dichos conjuntos de datos se obtuvo un 100% de sensibilidad, 81,6% de exactitud, 68,2% de especificidad y 91,5% de AUROC. Por otro lado, cuando se evaluó el modelo en un conjunto de 51 imágenes obtenidas de muestras retrospectivas de un laboratorio independiente de anatomía patológica, se obtuvo un 90,9% de sensibilidad, 62,7% de exactitud, 41,4% de especificidad y 74,1% de AUROC. Además se desarrolló un prototipo de interfaz gráfica para que los profesionales afines de diversas instituciones médicas del país puedan acceder a esta herramienta y eficientizar el proceso de elección del tratamiento óptimo.

En conclusión y tras un año de dedicación al desarrollo de nuestra plataforma digital, consideramos que proporcionar a pacientes con cáncer la capacidad de ser categorizados, inicialmente, en un subgrupo molecular mediante una coloración tan básica como la Hematoxilina-Eosina (H&E), representa una contribución significativa al avance en el diagnóstico de *drivers* oncogénicos. Reconociendo que requerirá de un mayor número de casos y de mejoras en el modelo para perfeccionar los resultados, estamos convencidas de haber iniciado un camino que agilizará los tiempos de diagnóstico y ampliará el alcance de la detección de inestabilidad microsatelital y otros biomarcadores moleculares en cáncer.

Introducción

El cáncer de endometrio es el sexto cáncer diagnosticado con mayor frecuencia en mujeres y la neoplasia maligna ginecológica con mayor incidencia en países de ingresos altos [2]. La mayoría de estos casos son esporádicos, pero se estima que el 5% ocurre en el contexto de una patología hereditaria, más comúnmente en el Síndrome de Lynch, a partir de mutaciones en el sistema de reparación de errores del ADN (MMR). Dependiendo el tipo de tumor endometrial, el tratamiento puede ser más agresivo o más conservador, por lo que resulta de suma importancia identificar a las pacientes con mayor riesgo a la recurrencia y muerte, para garantizar un tratamiento acorde a su enfermedad, evitando daños y costos innecesarios. Los subgrupos moleculares, que permiten identificar pacientes con valor pronóstico (sobrevida) y predictivo (tratamiento) se definen mediante la determinación de distintos biomarcadores que en el cáncer de endometrio son: la proteína P53, el gen Pol-E y la inestabilidad microsatelital.

La inestabilidad microsatelital en particular, es una consecuencia de defectos en el proceso del sistema de reparación del ADN o MMR. Este es un mecanismo de reparación de errores de replicación del ADN a través de cuatro proteínas esenciales: MLH1, MSH2, MSH6 y PMS2. Cuando este mecanismo falla, el sistema de reparación es deficiente (*deficient* MMR o MMRd); por el contrario, cuando las cuatro proteínas están intactas y conservadas, el sistema es proficiente (*proficient* MMR o MMRp). En la actualidad, el tamizaje para determinar si un tumor es MMRp o MMRd, es efectuar inmunohistoquímica (IHQ) en búsqueda de la presencia de las cuatro proteínas. Debido a que esta técnica resulta un excelente subrogante de inestabilidad microsatelital, podemos separar a los pacientes con baja probabilidad de aquellos que tengan alta probabilidad de presentar este biomarcador. Todos los casos no concluyentes por IHQ e incluso los que resultan MMRd, deben continuar su estudio con otras técnicas de biología molecular para confirmar los distintos eventos moleculares. Las pruebas son muy costosas y requieren de un patólogo entrenado, biólogos y de un laboratorio preparado para efectuar estos diagnósticos, por lo que

resulta esencial simplificar tecnologías y ahorrar en tiempo y recursos. Cuando se habla de pacientes con cáncer, el tiempo es un recurso escaso y esencial y este tipo de pruebas no solo son costosas, sino que suelen retrasar el diagnóstico final.

Dado que la inestabilidad microsatelital es uno de los *drivers* oncogénicos más importantes, tanto en la patología oncológica en general como específicamente en el tumor de endometrio, es crucial destacar que si el tumor resulta ser inestable (MSI), el enfoque del tratamiento y el pronóstico para la paciente cambiarán significativamente, beneficiando en distinta medida según el estadio clínico. Este cambio hacia la medicina de precisión subraya la importancia de considerar el estado MSI para una atención médica más personalizada y efectiva.

2.1. Objetivos del trabajo

Con el fin de simplificar la metodología de diagnóstico de MSI, la propuesta de este trabajo es el desarrollo de una herramienta de patología digital para la clasificación de carcinomas endometriales proficientes y deficientes para el sistema de reparación del ADN, mediante un algoritmo de inteligencia artificial. El objetivo además, es crear una herramienta que acorte el tiempo al diagnóstico, ayude en la toma de decisiones del patólogo y permita invertir o no, luego de este tamizaje, en análisis moleculares. El flujo de trabajo consiste en obtener imágenes tumorales (WSI) digitalizadas de las cuáles ya se sepa su estado microsatelital. Estas imágenes se utilizan para entrenar a un modelo de aprendizaje profundo para que luego pueda ser capaz de predecir la estabilidad o inestabilidad microsatelital de una imagen desconocida. Adicionalmente, se busca armar un prototipo de interfaz gráfica para que laboratorios de anatomía patológica del país puedan acceder a esta herramienta.

Además del objetivo impulsor del trabajo, la meta principal es desarrollar un modelo con potencial para ser implementado como parte integral de los procesos rutinarios en entornos médicos, con una sensibilidad que supere el 90%. Se aspira alcanzar un hito ambicioso: aprender del modelo. Si el modelo adquiere el desempeño deseado, se le podría solicitar que identifique aquellos patrones que motivan las distintas clasificaciones. En lo que se refiere al presente trabajo, en la actualidad, la inestabilidad microsatelital es un fenómeno molecular imperceptible para el ojo humano. Si el algoritmo lograra identificar patrones y clasificar correctamente a las imágenes digitalizadas como MSI, tales patrones podrían ser examinados por

un patólogo experto y se podría establecer una relación entre dichos patrones y sus respectivas características tisulares. Este análisis, contribuiría significativamente a la ciencia médica, al asociar las características tisulares identificadas en las imágenes digitalizadas con la presencia del biomarcador. Se complementarían herramientas de laboratorio complejas como la Inmunohistoquímica o la PCR constituyendo un complemento rápido y costo-efectivo en el diagnóstico. Esta colaboración entre la IA y los profesionales de salud podría revolucionar la forma en que se aborda el diagnóstico médico, mejorando así la atención y el tratamiento de los pacientes.

El desarrollo es de especial utilidad en un país como Argentina, donde los recursos de alto costo y los profesionales entrenados no están disponibles en todas las regiones para efectuar estas determinaciones. Puede resultar también, en mayor información para el oncólogo y una mejora en la toma de decisiones para las pacientes con cáncer de endometrio, que tendrán un tratamiento dirigido y acorde a su enfermedad.

Para concluir, esperamos que con el advenimiento del procesamiento de imágenes y de la IA orientada a la medicina, esta herramienta pueda ser utilizada en un laboratorio de rutina de Anatomía Patológica y afines. La aplicación se realizaría a partir del corte histológico coloreado con Hematoxilina-Eosina del tumor de un paciente. El objetivo sería identificar la inestabilidad microsatelital, actuando como un sistema de soporte en la toma de decisiones, es decir, una variable más a la hora de determinar el plan de tratamiento.

Queremos resaltar además que, basándonos en nuestro algoritmo, podrían desarrollarse otros con la misma o una similar metodología de inteligencia artificial. Éstos estarían dirigidos a distintos biomarcadores en este u otro tipo de cáncer, ofreciendo un aporte significativo en futuras investigaciones clínicas.

Marco teórico

3.1. Cáncer de endometrio

El endometrio es el tejido que recubre el útero, un órgano hueco, muscular con forma de pera invertida, en donde crece y se desarrolla el feto durante el embarazo. Se encuentra en la pelvis femenina por delante del recto, detrás de la vejiga y en contacto con la vagina. Las trompas de falopio entran por ambos lados en su parte superior [3]. El útero se compone de dos partes principales: la primera, conocida como el cuerpo del útero o Corpus Uteri, constituye la porción superior del órgano; la segunda, que conecta el útero con la vagina, se denomina cuello del útero o cérvix. Si bien existe también el cáncer cervical (Cervix Uteri), en este trabajo cuando se hable de cáncer de útero se estará haciendo referencia al cáncer de cuerpo uterino (Corpus Uteri) [4]. La anatomía de dicho órgano puede observarse en la siguiente figura:

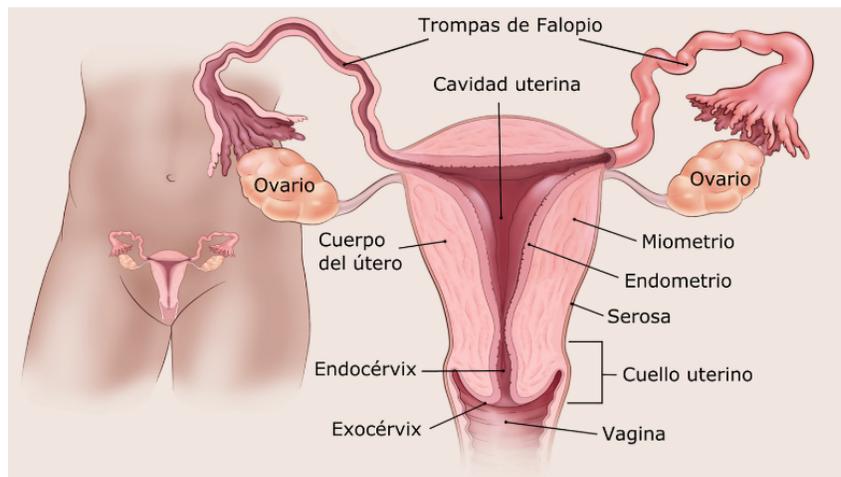


Figura 3.1: Anatomía del útero [4]

Tal como se puede ver en la Figura 3.1, el cuerpo del útero tiene tres partes

principales, el miometrio, el endometrio y la serosa o perimetrio. La capa media es el miometrio, una capa gruesa muscular que se contrae involuntariamente de acuerdo a las necesidades del sistema reproductor como la menstruación, el embarazo o el parto. Por otro lado, el endometrio es la capa interna. Durante el ciclo menstrual femenino, la morfología del endometrio va variando de acuerdo a las hormonas presentes en las distintas fases. Los niveles hormonales regulan el crecimiento y vascularización del endometrio para la implantación, seguido de la menstruación si no hay embarazo, repitiéndose hasta la menopausia [5]. Por último, la capa externa es la serosa o el perimetrio, y sus principales funciones son la protección y la lubricación [6]. La siguiente figura muestra la histología del útero identificando claramente sus tres capas distintas.

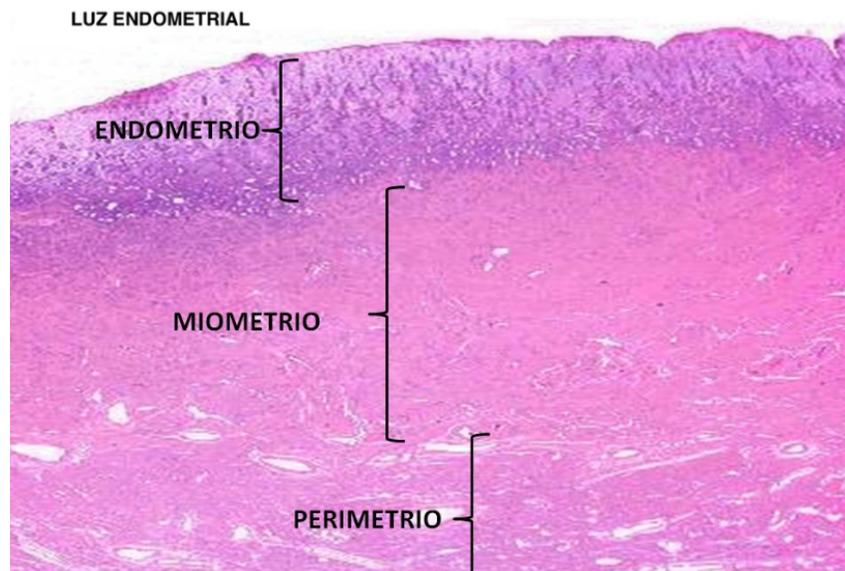


Figura 3.2: Histología del útero. Endometrio, miometrio y perimetrio [7]

Según la organización mundial de la salud (OMS), el cáncer de cuerpo de útero, histológicamente, puede referirse a los tumores epiteliales del endometrio, pero también, a otros tipos tumorales como las neoplasias mesenquimales y dentro de éstos al sarcoma uterino, los tumores mixtos, etc. En la Figura 9.1 del anexo se encuentra la clasificación más detallada de la OMS.

Sin embargo, el carcinoma endometrial representa alrededor del 95% de todos los casos de esta neoplasia, en tanto que el diagnóstico de sarcoma uterino es poco

frecuente [8]. Los sarcomas se desarrollan a partir del estroma uterino y el carcinoma endometrial a partir del epitelio del revestimiento interno uterino. [3].

Además del cáncer uterino existen también otras lesiones benignas como miomas y endometriosis. Los miomas son tumores benignos que se generan en el miometrio y la endometriosis es la aparición y crecimiento de tejido endometrial fuera del útero aumentando el riesgo para el desarrollo de otros tumores [3].

3.1.1. Incidencia y mortalidad

Como se mencionó previamente, el cáncer uterino es el sexto cáncer más frecuente en mujeres en todo el mundo con un estimado de 417.367 nuevos casos diagnosticados y 97.370 muertes en el año 2020. Esto corresponde a una incidencia de $8,7^1$ y a una mortalidad de 1,8 por cada 100.000 mujeres en el mundo [9]. Es el cáncer ginecológico más frecuente en países desarrollados. De hecho su incidencia es ampliamente mayor en dichos países que en países en desarrollo. La siguiente figura muestra la comparación entre las tasas de incidencia de dicho cáncer en los distintos países del mundo.

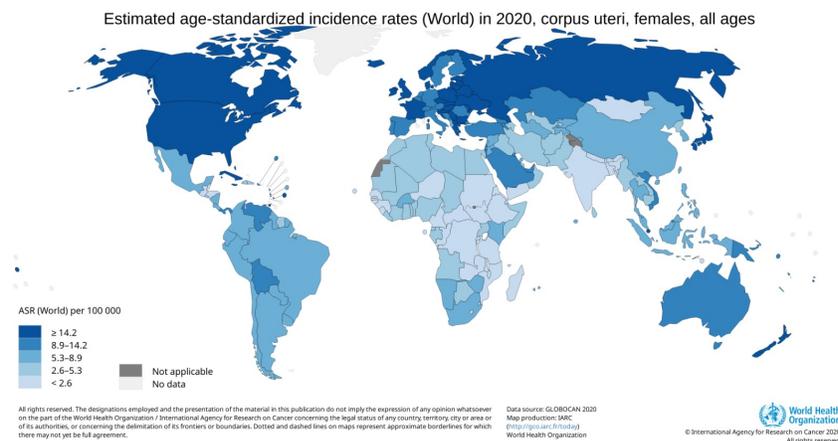


Figura 3.3: Tasas de incidencia de cáncer de cuerpo de útero estandarizadas por edad estimadas a nivel mundial en el año 2020 para todas las mujeres de todas las edades [9]

Por otro lado, en Argentina, la incidencia es de 7,6 y la mortalidad de 2,3. Si se

¹Se utiliza una tasa estandarizada por edad (ASR, Average Standard Rate) ya que es una medida resumida de la tasa que se habría observado si la población tuviera una estructura de edad estándar. La estandarización es necesaria cuando se comparan varias poblaciones que difieren con respecto a la edad ya que la misma tiene una fuerte influencia en el riesgo de cáncer.

analiza el caso de los Estados Unidos, dicho cáncer tiene una incidencia de 21,4 y una mortalidad de 3,1 [9]. La siguiente figura muestra la comparación entre ambos países:

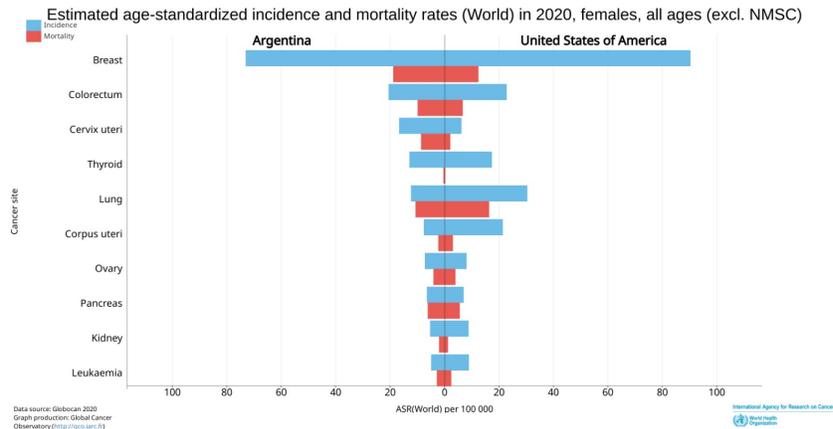


Figura 3.4: Tasas de incidencia y de mortalidad de los 10 cánceres más frecuentes estandarizadas por edad estimadas a nivel mundial en 2020 para todas las mujeres de todas las edades. Comparación entre Argentina y Estados Unidos [9]

La mediana de edad al diagnóstico del cáncer de endometrio en Argentina es 65 años y su tasa de supervivencia a 5 años es del 80 % [10]. Estos números son muy similares a los de Estados Unidos que tienen una mediana de edad de 60 años y una tasa de supervivencia a 5 años del 81 % [11].

3.1.2. Factores de riesgo

En primer lugar, un exceso a la exposición a estrógenos es considerado un gran factor de riesgo. Tal como se mencionó con anterioridad, esto se debe a que el estrógeno estimula el crecimiento del endometrio. Un crecimiento descontrolado puede derivar en cáncer. Dicho exceso puede ser atribuido tanto a causas endógenas como exógenas. Dentro de las causas endógenas, las más comunes son la obesidad, un tumor productor de estrógeno, ciclos anovulatorios, menarquía precoz y menopausia tardía. En cuanto a las causas exógenas, las más comunes son la terapia estrogénica con y sin progestágenos, la terapia con tamoxifeno y los fitoestrógenos. Por otro lado, los factores de riesgo no hormonales más significativos son la edad, los antecedentes familiares, las mutaciones genéticas, de las cuales se hablará en más profundidad, la infertilidad, la etnia (las mujeres caucásicas tienen un mayor

riesgo de desarrollar cáncer de endometrio que las mujeres de otras etnias), la región geográfica y radioterapia pélvica previa por otra neoplasia [3].

3.1.3. Carcinoma de endometrio hereditario

El cáncer de endometrio puede ser esporádico o hereditario. Estos últimos corresponden al 2-5 % de todos los cánceres de endometrio. Si bien el riesgo poblacional de que una mujer desarrolle este tipo de cáncer es bajo, la presencia de antecedentes familiares o condiciones hereditarias puede aumentar significativamente este riesgo, llegando hasta un 70 %. Los síndromes hereditarios más frecuentemente relacionados con el cáncer de endometrio son el síndrome de Lynch y el síndrome de Cowden. En los últimos años, también se ha relacionado el síndrome de cáncer de mama y de ovario hereditario (BRCA1/2) con el desarrollo de cáncer de endometrio. Otros síndromes hereditarios posiblemente relacionados son el síndrome de Peutz-Jeghers y el síndrome de Li-Fraumeni. De todos estos síndromes, el más común es el síndrome de Lynch que surge de un defecto en el sistema de reparación del ADN [3].

Carcinoma de endometrio y Síndrome de Lynch

Esta patología, el síndrome de Lynch, merece una breve consideración aparte. Se trata de un síndrome autosómico dominante de predisposición al cáncer, en el cual los afectados suelen heredar una variante inactivadora en uno de los cuatro genes del sistema de reparación de errores del ADN: MLH1, MSH2, MSH6 o PMS2. Este sistema, altamente conservado, es responsable de corregir los errores de inserción y eliminación que se producen durante la replicación genómica. La pérdida de la función del MMR conduce a la inestabilidad de los microsatélites, generando un fenotipo hipermutado y una mayor susceptibilidad al cáncer. Aunque los pacientes con síndrome de Lynch tienen un mayor riesgo de padecer diversas neoplasias malignas, es más común que desarrollen cáncer colorrectal y de endometrio. Como resultado, estos pacientes suelen tener una esperanza de vida más reducida en comparación con aquellas personas no afectadas [12].

3.1.4. Clasificación del cáncer de endometrio

Originalmente, la clasificación del carcinoma endometrial se basaba únicamente en criterios histológicos como los que se mencionaron previamente. Con el avance

de las técnicas de biología molecular, se introdujo también una clasificación molecular. El carcinoma endometrial, o adenocarcinoma, es la neoplasia más común en el útero y se origina en células glandulares. Los adenocarcinomas pueden desarrollarse en diversos tejidos, incluyendo pulmón, colon, páncreas, estómago y otros. En el adenocarcinoma de endometrio, las células glandulares experimentan un crecimiento anormal y una proliferación descontrolada, pudiendo invadir el tejido uterino circundante y extenderse fuera del órgano [13].

En referencia a los carcinomas endometriales, la clasificación de la FIGO (Federación Internacional de Ginecología y Obstetricia) divide los tumores en dos tipos principales. El tipo I, también conocido como adenocarcinomas endometrioides grado 1 y 2, constituye el grupo más prevalente (70-80 %) y presenta un bajo potencial maligno. Estos tumores son asociados con estrógeno, vinculados a la obesidad, y muestran un pronóstico favorable. Por otro lado, los tumores de tipo II, que incluyen adenocarcinomas endometrioides grado 3, serosos y otros subtipos histológicos, son menos frecuentes (20-30 %) y de alto grado. Estos tumores son biológicamente agresivos, más comunes en mujeres con peso saludable y no responden a la estimulación de estrógeno [3].

Clasificación molecular

Esta clasificación sirve para estratificar a los tumores epiteliales según número de mutaciones y es de suma utilidad para predecir el pronóstico o la sensibilidad a inmunoterapia. Hoy en día esta clasificación forma parte de las nuevas guías para el desarrollo del plan terapéutico del paciente. Esta clasificación puede observarse en la siguiente figura:

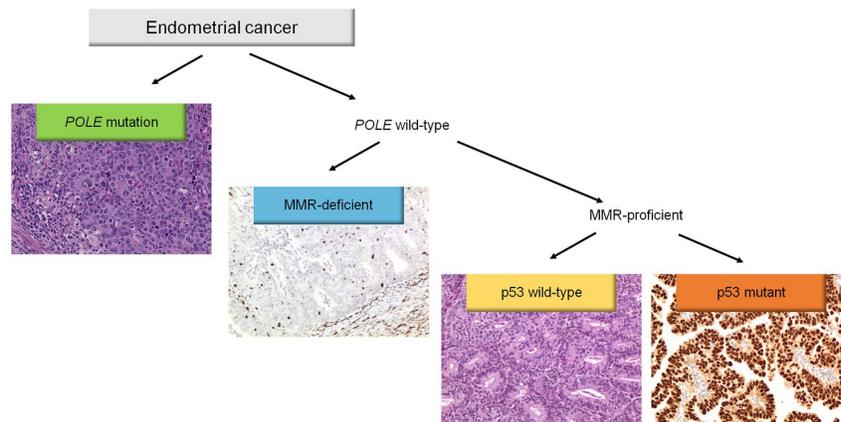


Figura 3.5: Clasificación molecular del cáncer de endometrio [14]

La clasificación molecular de los carcinomas endometriales se resume en cuatro categorías principales. Los tumores *POLE* mutados, representando el 5-15 %, exhiben un número muy elevado de mutaciones, morfología endometriode, alto infiltrado inmunitario y un muy buen pronóstico. Aquellos con MMRd, constituyendo el 20-30 %, muestran un número alto de mutaciones, son en su mayoría endometrioides, presentan mutaciones en genes MMR y tienen un pronóstico intermedio. Los tumores con alteraciones en el gen *p53*, comprendiendo el 10-25 %, exhiben elevada variación del número de copias, carga mutacional muy baja, y se presentan en histologías serosas, células claras, carcinosarcoma y algunos endometrioides, con un pronóstico desfavorable. Por último, la categoría de tumores no específicos de subtipos moleculares (NSMP) abarca el 30-60 %, caracterizándose por escasa variación del número de copias, carga mutacional baja, endometrioides de bajo grado o histología mixta, y pronóstico intermedio a malo [3].

3.1.5. Diagnóstico

El diagnóstico clínico consiste en la recopilación de información relevante, como la existencia de factores de riesgo, antecedentes familiares o condiciones hereditarias. Para confirmar la presencia de un tumor, se realizan exploraciones ginecológicas, ecografías transvaginales, histeroscopías y biopsias. Posteriormente, se llevan a cabo pruebas complementarias para identificar marcadores tumorales o para obtener imágenes detalladas del tumor, utilizando técnicas como resonancia magnética, tomografía computarizada o PET-CT [3].

El sangrado genital anómalo constituye el signo clínico crucial para diagnosticar

el cáncer de endometrio. El diagnóstico definitivo se lleva a cabo mediante el estudio histológico, que generalmente se realiza en coágulos provenientes de sangrado genital, biopsias por histeroscopia o piezas quirúrgicas. Este procedimiento permite la categorización precisa de la neoplasia y facilita la realización de estudios moleculares específicos en relación con la patología de cada paciente [3].

3.2. Inestabilidad microsatelital y cáncer de endometrio

3.2.1. Microsatélites

En genómica, un microsatélite es un fragmento corto y repetido de ADN. Suelen estar formados por patrones de no más de 6 pares de bases, los cuales se repiten hasta 50 veces. Estos fragmentos se encuentran en todo el genoma, tanto en las partes codificantes como en las no codificantes [15]. En la Figura 3.6 se puede observar un ejemplo de un microsatélite dinucleótido, es decir que repite un patrón de dos pares de bases.

```
CGTATGCTAGTCTGGTACGACGTATCGATCGATTGCATAGCTACTCTCTCTCT
CTCTCTCTCTCTCTCTCTCTCTCTCGATCAGCTAGTAGCTAACGAGCGATC
GATTGCATCGACTACTAGCTA
```

Figura 3.6: Microsatélite dinucleótido

La presencia de los microsatélites tiene una gran incumbencia en el cáncer por la forma en que éstos mutan. Al igual que todo el ADN, cuando los microsatélites son replicados, se pueden generar errores, encontrando en la nueva cadena bases de más, de menos o incorrectas. Se cree que la mayor cantidad de mutaciones en los microsatélites se da por deslizamiento durante la replicación [16]. En este tipo de mutaciones, la ADN polimerasa, enzima responsable de ubicar las nuevas bases en la cadena complementaria de ADN siendo sintetizada, se desasocia de la cadena molde, y se vuelve a asociar en el lugar incorrecto. Esta disociación puede generar dos tipos de mutaciones. En primer lugar, es posible que la enzima se saltee una secuencia repetitiva, causando que la cadena complementaria sea más corta que la cadena molde. Por otro lado, puede suceder que la polimerasa se deslice hacia atrás, sintetizando el mismo segmento dos veces, generando una cadena complementaria más larga que la cadena molde [17]. En la Figura 3.7 se puede observar la replicación normal así como ambos casos de deslizamientos mencionados.

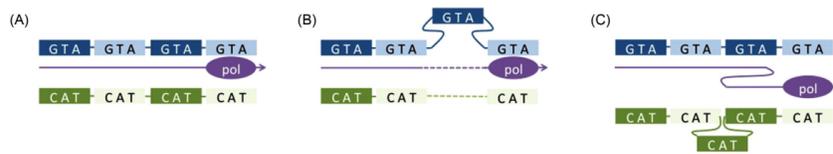


Figura 3.7: Replicación normal de un microsatélite (A), deslizamiento hacia adelante (B) y deslizamiento hacia atrás (C) [17]

3.2.2. Mecanismo de reparación del ADN

Dado que los errores de replicación ocurren constantemente, el ser humano cuenta con mecanismos de reparación para garantizar que el ADN se replique de forma correcta y que las mutaciones no perduren. El mecanismo de *mismatch repair* (MMR) es un sistema de reparación del ADN que puede identificar y corregir errores de replicación en la cadena recién sintetizada [18]. El MMR logra identificar errores cuando las bases fueron apareadas de forma incorrecta, así como la existencia de inserciones o deleciones de bases. Este sistema reconoce el error, remueve el pedazo de la cadena nueva que contiene dicho error, y llena el agujero con las bases correctas.

Diversos estudios llevados a cabo en bacterias como *Escherichia Coli* han demostrado que al silenciar ciertos genes, el ADN muta a niveles mucho más altos de los usuales. Estos genes codifican las proteínas Mut, las cuales se ven altamente conservadas en seres humanos. En el caso de las bacterias, tres de estas proteínas son esenciales para el mecanismo MMR, llamadas MutS, MutH y MutL. Las proteínas homólogas a la MutS en humanos son la MSH2 y la MSH6. Juntas forman un heterodímero que puede reconocer los errores de la cadena recientemente sintetizada [19].

Por otro lado, la proteína MutL tiene muchos homólogos en humanos, pero la proteína más importante para el mecanismo MMR es el heterodímero de MLH1 y PMS2. Éstas cumplen el importante rol de estimular e interactuar con otras proteínas, como las exonucleasas o helicasa para eliminar las bases incorrectas de la cadena de ADN. De esta forma, la ADN polimerasa puede intervenir y ubicar las bases correctas en los lugares correspondientes. En la Figura 3.8 se puede observar una esquematización del mecanismo MMR y cómo intervienen las proteínas mencionadas.

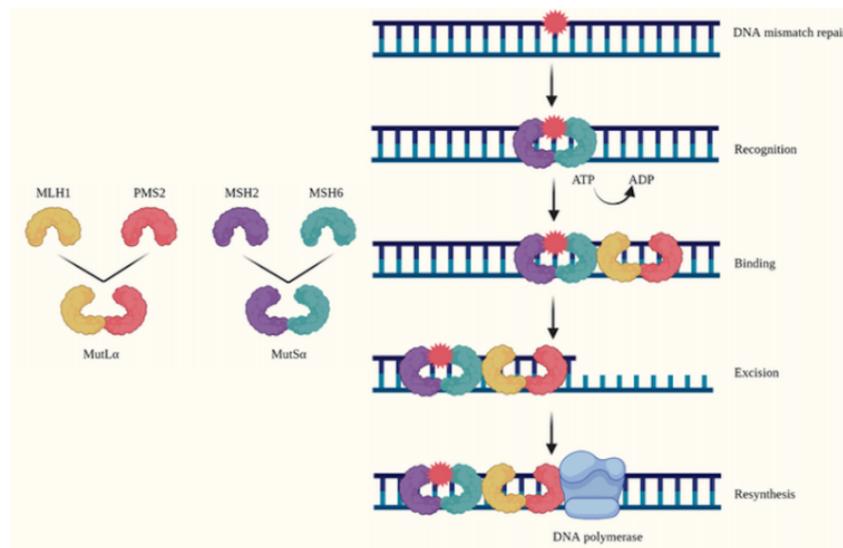


Figura 3.8: Mecanismo MMR y las proteínas involucradas [20]

De esta manera, el organismo logra eliminar los errores generados durante el proceso de replicación. Por la naturaleza repetitiva de los microsatélites, los errores de deslizamiento son muy comunes. Por lo tanto, el sistema de reparación es de suma importancia para mantener la estructura correcta de estos sectores del ADN.

3.2.3. Inestabilidad microsatelital y sus variantes somática y germinal

En el caso de que el sistema de reparación MMR se encuentre dañado, los errores no van a ser reparados. Por ende, cuando la cadena complementaria se convierte en el molde para una replicación posterior, la mutación persiste a lo largo de las sucesivas replications. De esta manera, dependiendo de la naturaleza de la mutación, el nuevo microsatélite es más corto o más largo que en la cadena original. Estos casos son conocidos como inestabilidad microsatelital.

Es importante destacar que la inestabilidad microsatelital y un mecanismo MMR deficiente no son equivalentes, ya que se trata de fenómenos biológicos distintos. Dado que hay una gran concordancia entre estos fenómenos, en este trabajo y en muchos otros se asume que al decir que un tumor tiene el mecanismo MMR deficiente, tiene inestabilidad microsatelital, y viceversa.

Dentro de la inestabilidad microsatelital, existen variantes somáticas y germinales. Como fue mencionado, esta distinción habla sobre el origen de la mutación, es decir, si la misma es causada por una condición genética o si fue adquirida a lo largo

de la vida del paciente. En el caso de la variante germinal, los genes responsables del mecanismo MMR están mutados en todas las células del paciente desde su nacimiento. Esta variante está asociada al síndrome de Lynch que fue explicado en la sección 3.1.3. Por otro lado, un paciente puede presentar un tumor con inestabilidad microsatelital sin tener una predisposición genética. En este escenario, la MSI es de tipo somático, siendo la más común entre las dos variantes [21]. La pérdida de la expresión de la proteína MLH1, asociada con la inestabilidad somática, se produce debido a la metilación del promotor de dicho gen. Este fenómeno epigenético implica la adición de grupos metilo al gen promotor de MLH1, responsable de regular la iniciación de la transcripción de este gen. En consecuencia, la hipermetilación del promotor causa la inhibición del gen y se refleja en la falta de producción de MLH1 [22].

3.2.4. Diagnóstico de inestabilidad microsatelital

Inmunohistoquímica

La técnica de diagnóstico más utilizada para identificar la inestabilidad microsatelital en un tumor es la inmunohistoquímica. Este método observa el funcionamiento del mecanismo MMR, evaluando la pérdida de expresión de una o más de las cuatro proteínas clave de este mecanismo: MLH1, MSH2, MSH6 y PMS2. De esta manera, actúa como un subrogante eficaz para detectar la inestabilidad microsatelital en el tumor.

Esta técnica tiene el fin de hacer visible un antígeno, generalmente una proteína, de la biopsia de un tumor para que un médico patólogo pueda determinar su presencia o ausencia. Utiliza la unión específica entre un antígeno y un anticuerpo para lograr su visualización [23]. La misma puede ser directa o indirecta. En la inmunohistoquímica directa se utiliza un solo anticuerpo marcado para detectar directamente la proteína de interés en la muestra de tejido, mientras que en la indirecta se emplean dos anticuerpos: uno específico para la proteína y otro secundario marcado, amplificando la señal. Ambos métodos son cruciales en el diagnóstico e investigación biomédica [24].

En el laboratorio, el procedimiento implica teñir una muestra con hematoxilina y eosina para clasificar la neoplasia, identificar sus características, evaluar la cantidad de tumor en cada portaobjetos y asegurarse de que el tejido esté bien fijado y sea viable. Después de verificar la idoneidad de la muestra, se realiza la incubación con el

anticuerpo anti-MLH1, anti-MSH2, anti-MSH6 y anti-PMS2 en cuatro portaobjetos diferentes, asignando cada proteína a un portaobjetos específico [25]. A continuación, la Figura 3.9 muestra una esquematización del proceso.

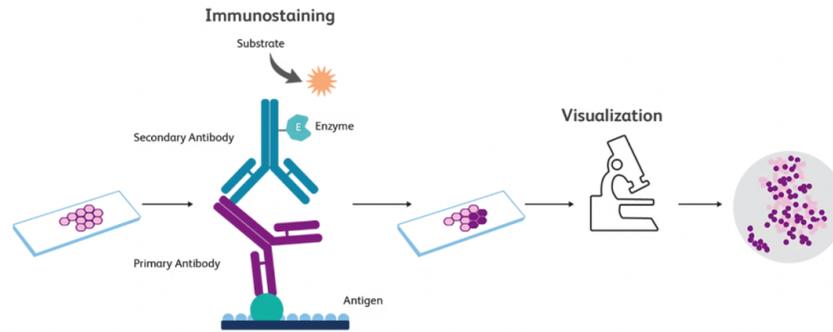


Figura 3.9: Flujo de trabajo del proceso de inmunohistoquímica (ejemplificado con inmunohistoquímica indirecta) [26]

Las muestras deben ser evaluadas bajo un microscopio por un patólogo entrenado para detectar la presencia o ausencia de cada una de las proteínas. Para que el patólogo pueda determinar que el mecanismo MMR se encuentra intacto, debe haber tinción de los núcleos de las células tumorales para las cuatro proteínas. En el caso de que una de ellas no tenga tinción del núcleo, se puede concluir que no hay presencia de la proteína en cuestión y que por lo tanto, el mecanismo MMR se encuentra dañado. En la Figura 3.10 se pueden observar ejemplos de una tinción positiva (izquierda) y negativa (derecha) para la proteína MSH2 en los núcleos tumorales.

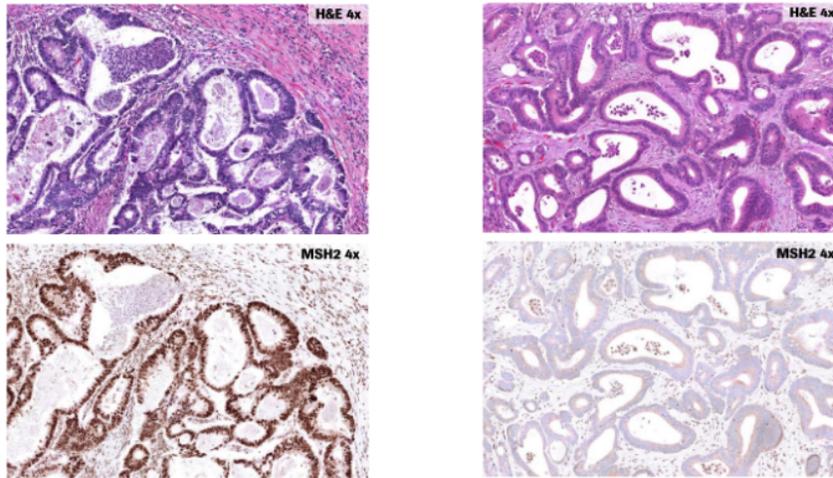


Figura 3.10: Proteína MSH2 presente (izquierda) y ausente (derecha) en núcleos tumorales mediante IHQ [25]

Si el mecanismo MMR resulta deficiente, se determina que el tumor del paciente presenta inestabilidad microsatelital. No obstante, en muchos laboratorios, se realiza una segunda prueba de biología molecular para confirmar este diagnóstico y poder seleccionar el tratamiento adecuado. Además, en situaciones en las que la conclusión por inmunohistoquímica no sea clara para alguna de las proteínas, también se recomienda recurrir a pruebas de biología molecular.

Biología molecular

La detección directa de inestabilidad microsatelital se realiza a través de pruebas de reacción en cadena de la polimerasa (PCR). Esta técnica amplifica millones de copias de un segmento de ADN, específicamente del microsatélite en cuestión. Posteriormente, se emplea la electroforesis en gel para visualizar los resultados, permitiendo la comparación del microsatélite con un marcador de tamaño conocido. Si se observa una discrepancia en longitud, ya sea que el microsatélite sea más largo o más corto de lo esperado, se concluye que existe inestabilidad [27].

Se repite este proceso para múltiples microsatélites en el tejido tumoral. El Instituto Nacional del Cáncer (NCI) de Estados Unidos recomienda analizar cinco marcadores (BAT-25, BAT-26, D2S123, D5S346 y D17S250) como indicadores microsatelitales para determinar el estado microsatelital de un tumor [28]. Si los cinco marcadores muestran estabilidad, la muestra se clasifica como MSS. En cambio, si solo un marcador presenta inestabilidad, se clasifica como MSI-L, y si dos o más

marcadores son inestables, se clasifica como MSI-H. Es importante destacar que en los casos MSI-L, la IHQ suele resultar indeterminada.

De esta manera, se dispone de dos técnicas de laboratorio: una para evaluar la integridad del mecanismo MMR (IHQ) y otra para visualizar directamente la estabilidad o inestabilidad del microsatélite (PCR). Diversos estudios comparativos entre estas metodologías han revelado disparidades en el diagnóstico, oscilando entre el 2% y el 8% de los casos en diversas instituciones [29]. Dada la baja discrepancia, se consideran aceptables los resultados obtenidos por cualquiera de las dos técnicas, aplicándose ambas solo en casos no concluyentes o cuando la muestra resulta inestable por IHQ. Sin embargo, es crucial señalar que esta discrepancia existe y puede ser origen de posibles errores.

3.2.5. Inestabilidad microsatelital y su mejor pronóstico

Se establece una clasificación del cáncer que distingue entre tumores "fríos" y "calientes". Los tumores con inestabilidad microsatelital presentan elevados niveles de mutaciones, facilitando la generación de neoantígenos, es decir, antígenos tumorales capaces de inducir una respuesta inmune por parte del huésped. Esta capacidad de producir neoantígenos clasifica al tumor como "caliente" [30]. En la Figura 3.11, se presenta una representación esquemática que ilustra las diferencias entre un tumor "frío" y uno "caliente", destacando la notable presencia de células del sistema inmune en el tumor "caliente".

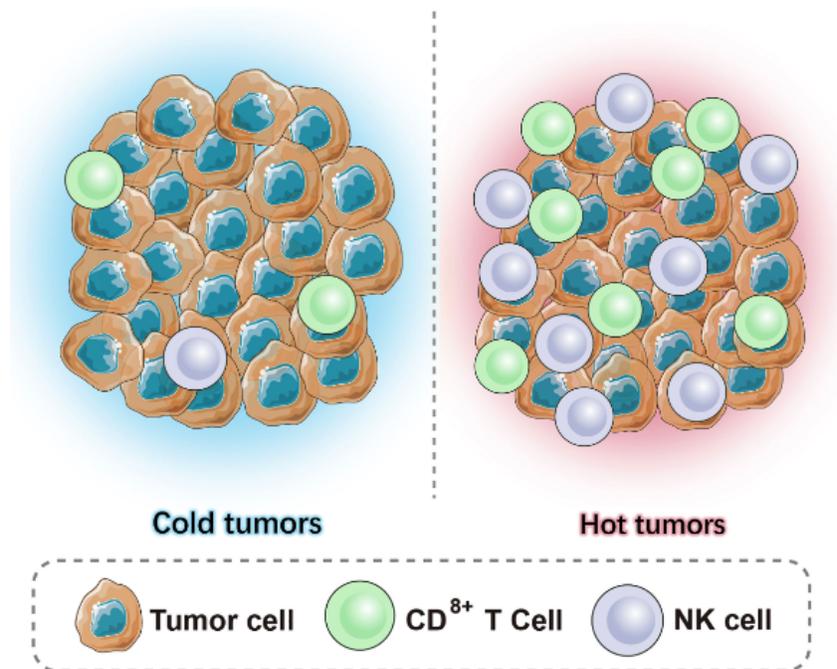


Figura 3.11: Tumor "frío" y tumor "caliente" [31]

Existen diversos factores que determinan si un tumor es "frío" o "caliente", entre ellos se incluyen checkpoints inmunológicos, la presencia de linfocitos infiltrantes de tumores, el tejido linfoide terciario, mutaciones específicas, así como la inestabilidad microsatelital [30].

Históricamente, las pruebas de inestabilidad microsatelital para mujeres con cáncer de endometrio sólo se llevaban a cabo en el caso de que la paciente tuviera sospecha de síndrome de Lynch [32]. Hoy, la clasificación de un tumor como "frío" o "caliente" desempeña un papel crucial en la formulación de tratamientos personalizados para las pacientes afectadas. Los tumores "fríos", al carecer de la producción de neoantígenos, no son identificados eficazmente por el sistema inmune, lo que limita la eficacia de la inmunoterapia en estos casos. En contraste, los tumores "calientes" son susceptibles a la inmunoterapia, una estrategia de tratamiento que ofrece pronósticos más favorables que la quimioterapia o radioterapia convencional, mejorando así la calidad de vida durante el tratamiento [33]. Esta categorización permite optimizar la atención, adaptando los enfoques terapéuticos según las características específicas de cada tumor, lo que se traduce en mejores perspectivas de erradicación y una mejora general en el bienestar del paciente.

3.3. Tratamiento

El principal desafío que enfrenta el tratamiento del cáncer de endometrio radica en su agresividad. En la mayoría de los casos, la intervención implica procedimientos quirúrgicos que, lamentablemente, pueden privar a las mujeres de su capacidad de ser madres si así lo desean, una preocupación particularmente significativa para aquellas que aún se encuentran en edad fértil. Ésta es la razón por la cual se están llevando a cabo constantemente investigaciones y ensayos clínicos con el objetivo de desarrollar terapias menos invasivas. La clave reside en identificar con precisión los casos de cáncer de mayor gravedad y aquellos con menor gravedad, a fin de evitar que todas las pacientes reciban tratamientos sumamente agresivos e innecesarios. La correcta clasificación del tipo de tumor es esencial para ofrecer el tratamiento más adecuado. En la actualidad, una alternativa terapéutica que brinda un pronóstico prometedor es la inmunoterapia, una opción de tratamiento que se explorará en detalle más adelante, analizando su aplicación y cuáles son los casos en donde muestra mayor efectividad.

Como se mencionó previamente, el tratamiento del cáncer de endometrio consiste en la mayoría de los casos en una intervención quirúrgica. El tratamiento varía de paciente a paciente ya que según el estadio clínico del tumor se establece la magnitud de la intervención. Existe incluso la posibilidad de participar en algún ensayo clínico en donde se esté probando una nueva alternativa de tratamiento. A continuación se mencionan brevemente los tratamientos más frecuentes. Se utilizan cinco tipos de tratamiento estándar: la cirugía, la quimioterapia, la radioterapia, la terapia con hormonas y la terapia dirigida.

La **cirugía** consiste en la extirpación del cáncer mediante un tratamiento quirúrgico y es la terapia más frecuente para el cáncer de endometrio. Según aumenta el estadio del cáncer, aumenta la agresividad de la intervención. Una vez que se extirpa todo el tumor visible, es posible que la paciente reciba alguna terapia adyuvante para terminar de eliminar las células cancerosas que pueden permanecer luego de la cirugía. Nuevamente, el tipo y agresividad de dicha terapia también dependen del grado del tumor. A continuación se explican brevemente las terapias adyuvantes más frecuentes [13].

La **radioterapia** es un enfoque de tratamiento del cáncer que emplea rayos-X de alta energía o diversas formas de radiación con el propósito de destruir células

cancerosas o prevenir su proliferación. Se puede tratar de radioterapia externa o interna. No solo se utiliza para tratar el cáncer sino también como terapia paliativa para calmar los síntomas y mejorar la calidad de vida [13].

Por otro lado, la **quimioterapia** es un tratamiento contra el cáncer que utiliza medicamentos para eliminar las células cancerosas. Estos medicamentos pueden administrarse por vía oral, intravenosa o directamente en el tumor. En el cáncer de endometrio, la quimioterapia se emplea para tratar la enfermedad que se ha propagado o que ha reaparecido tras un tratamiento. También puede utilizarse después de la cirugía para prevenir la reaparición del cáncer. Los medicamentos de quimioterapia tienen diversos mecanismos de acción, algunos dañan el ADN de las células cancerosas, lo que detiene su crecimiento y división, mientras que otros bloquean el suministro de nutrientes esenciales para las células cancerosas. Los efectos secundarios de la quimioterapia varían según el medicamento y la dosis, pero suelen incluir fatiga, náuseas, vómitos, pérdida de apetito, caída del cabello, infecciones y trastornos sanguíneos [34].

Por último, la **terapia con hormonas** es un enfoque de tratamiento contra el cáncer que suprime o bloquea la influencia de las hormonas, frenando así el crecimiento de las células cancerosas. En algunos casos, ciertas hormonas pueden estimular el crecimiento de determinados tipos de cáncer. Cuando las pruebas indican que las células cancerosas tienen receptores que las hacen susceptibles a las hormonas, se utilizan medicamentos, cirugía o radioterapia para reducir la producción de las mismas o bloquear su función [13].

3.3.1. Terapia dirigida

Éste es un enfoque de tratamiento que emplea medicamentos y otras sustancias con el propósito de identificar y atacar células cancerosas específicas. En comparación con la quimioterapia o la radioterapia, estas terapias suelen causar menos daño a las células normales. En el tratamiento del cáncer de endometrio, se utilizan sustancias como los anticuerpos monoclonales, los inhibidores de mTOR y los inhibidores de la transducción de señales como ejemplos de terapias dirigidas [13].

La inmunoterapia representa un importante avance en oncología, ofreciendo una alternativa de tratamiento para tumores que responden poco a la quimioterapia, con la ventaja adicional de minimizar los efectos adversos en la calidad de vida de los pacientes en comparación con otros enfoques terapéuticos. En mayo del 2017 la FDA

(Administración de Alimentos y Medicamentos de Estados Unidos) aprobó el uso de una droga llamada Pembrolizumab, un inhibidor de *checkpoints* inmunológicos. Se trata de un anticuerpo monoclonal que impide que las células tumorales supriman al sistema inmune del paciente permitiendo que el mismo ataque y elimine a las células tumorales [35]. La droga fue aprobada para usar en todos los casos de tumores con inestabilidad microsatelital o deficiencia del mecanismo MMR, sin importar el sitio primario del tumor.

3.4. Inteligencia artificial y aprendizaje profundo

La inteligencia artificial es un conjunto de herramientas y habilidades que buscan replicar la inteligencia humana en procesos computacionales, creando un campo dentro de la ciencia de la computación que tiene como objetivo la resolución de problemas [36]. Su aplicación es diversa, abarcando desde el control de robots en la robótica hasta el reconocimiento de patrones en datos, el aprendizaje automático y el procesamiento del lenguaje natural. Los beneficios potenciales de la IA para la sociedad son significativos. Destaca la mejora de la eficiencia y la productividad mediante la automatización de tareas, respaldando la toma de decisiones informadas al analizar grandes conjuntos de datos. Además, ofrece la posibilidad de desarrollar productos y servicios innovadores que podrían mejorar la calidad de vida de las personas, como diagnósticos médicos precisos o asistencia a personas con discapacidad [37].

Una de las ramas más conocidas de la IA es el aprendizaje automático o *machine learning* (ML), que se centra en el desarrollo de algoritmos para que las computadoras adquieran la capacidad de aprender de los datos y mejorar su rendimiento con el tiempo [38]. Dentro de ML, existe el aprendizaje profundo, o *deep learning*, un tipo de aprendizaje que utiliza múltiples capas para extraer de manera automática la información relevante de los datos, para luego aprender y tomar decisiones [39].

3.4.1. Redes neuronales

Los algoritmos de aprendizaje profundo se basan en el uso de redes neuronales para emular las conexiones neuronales presentes en el cerebro humano. En estas redes, las cuales están compuestas por varias capas, se pueden encontrar nodos, que representan las neuronas artificiales. Estas neuronas están interconectadas mediante

pesos, que son valores ajustables que determinan la fuerza y la dirección de la conexión entre ellas. Durante el proceso de aprendizaje, estos pesos se van actualizando de manera iterativa a medida que la red es entrenada con datos. La red neuronal se estructura en tres secciones clave: la capa de entrada, por donde se introduce la información; las capas ocultas, donde se lleva a cabo el procesamiento principal mediante la interacción de neuronas y ajuste de pesos; y la capa de salida, que proporciona el resultado final del algoritmo [40]. Para una comprensión visual de la relación entre las neuronas y sus conexiones ponderadas, se presenta una simplificación gráfica en la siguiente figura.

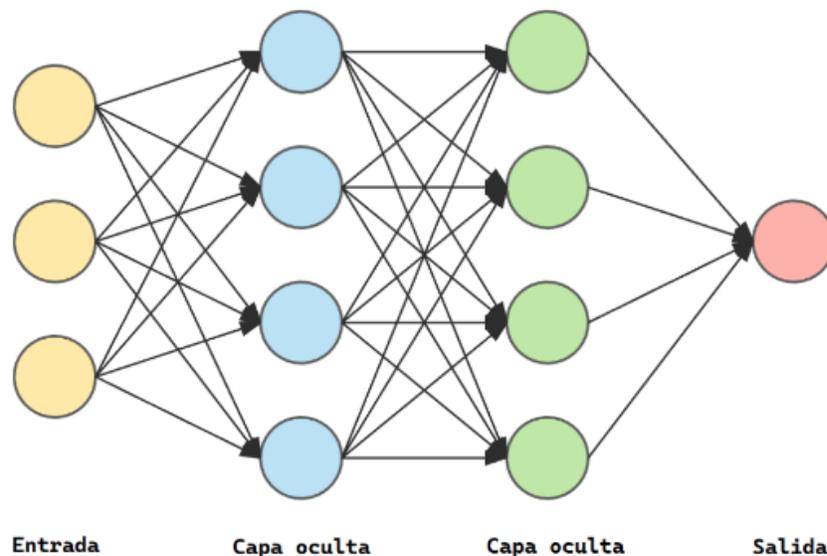


Figura 3.12: Esquema simplificado de una red neuronal

Dependiendo de la tarea que se busque realizar, se incluyen distintas capas en la red. Cada una de estas capas tiene un propósito específico que dicta cómo se relacionan sus neuronas de forma tal que la red adquiera conocimiento y capture patrones en los datos de forma efectiva. En particular, a continuación se mencionan las capas más utilizadas para el procesamiento de imágenes.

Capa convolucional

Una capa convolucional es una parte central de la arquitectura de una red neuronal convolucional principalmente utilizada para procesar imágenes. Al igual que en otras capas de una red, las conexiones entre neuronas se establecen mediante pesos

que se actualizan durante el proceso de aprendizaje. Estos pesos corresponden a los parámetros de los filtros utilizados en la capa convolucional. En este contexto, los filtros son ventanas pequeñas que se desplazan sobre la imagen de entrada, calculando la relación entre sus elementos y la imagen en cada posición (convolución). La aplicación de estos filtros genera mapas de activación, los cuales se apilan para formar el volumen de salida. La conexión de cada neurona en el mapa de activación a una región local de la entrada permite a la red aprender y reconocer patrones locales, aprovechando las correlaciones espaciales cercanas. Además, los parámetros del filtro se comparten en todas las posiciones locales, lo que reduce la cantidad de parámetros y contribuye a la eficiencia de la red. En términos de funcionalidad, las capas convolucionales iniciales se centran en capturar características de bajo nivel, como líneas y bordes, mientras que las capas posteriores se dedican a extraer características de alto nivel, como formas y objetos específicos [41].

Capa de *pooling*

En la capa de pooling, el objetivo es reducir la complejidad de la red neuronal mediante la agrupación de parámetros generados en capas anteriores. Esta reducción de dimensionalidad se logra a través de funciones de agregación, como el *max-pooling*, que selecciona el valor más alto de un grupo de parámetros, o el *average-pooling*, que calcula un valor promedio. La elección de la función depende de la tarea específica. Aunque esta capa implica pérdida de información, su función es simplificar significativamente el algoritmo [42].

Capa de atención

De manera análoga a cómo una persona enfoca su atención, en una red neuronal, una capa de atención permite que la red se concentre en partes específicas de la entrada [43]. Esto resulta especialmente beneficioso en tareas como el procesamiento del lenguaje natural, donde la red debe comprender las relaciones entre palabras o frases, así como en el procesamiento de imágenes para reconocer o segmentar objetos [44]. Las capas de atención, esencialmente, realizan una reducción ponderada de la media. En este proceso, se asigna un peso a cada elemento que contribuye a la media, determinado por una función de atención que considera las relaciones entre los distintos elementos de la entrada. Además, estas capas reducen la dimensión de la entrada al combinar los elementos en un único vector de salida [43].

Capa de regularización

En esta capa se busca resolver problemas de sobreajuste del algoritmo, que sucede cuando el modelo se torna muy personalizado para un conjunto de datos particular, y aprende de parámetros incorrectos, lo cual baja significativamente su desempeño para otros datos. El método *dropout* de regularización hace que, en cada iteración de entrenamiento, se ignore un porcentaje determinado de neuronas seleccionadas aleatoriamente antes de pasar a la siguiente capa. A este porcentaje se lo conoce como el *dropout rate*, y se lo determina antes del entrenamiento. Suele tomar un valor entre 0,2 y 0,5. Esto evita el sobreajuste, aumentando la habilidad de generalización del modelo [45].

Capa totalmente conectada

En este tipo de capas, también conocidas como capas densas, hay una interconexión total entre neuronas, es decir que todas las salidas de las neuronas de una capa son utilizadas como las entradas para todas las neuronas de la capa siguiente. Dado que otras capas no trabajan con la totalidad de la información, sino que de a grupos, estas capas son necesarias para lograr recopilar toda la información. Se suele contar con varias de estas capas, para bajar gradualmente el número de neuronas en la red [46]. De todos modos, estas capas son de alto costo computacional y no mantienen la estructura espacial de la entrada. Es por esto que estas capas suelen colocarse al final de la red y se utilizan principalmente para tareas de clasificación. En el caso de dichas tareas en donde la capa de salida de la red es una capa totalmente conectada, como función de activación, es decir, la función que transforma las entradas en salidas, se suele utilizar una función *softmax*. Esta función convierte un vector cuya dimensión coincide con el número de clases posibles para la clasificación en otro vector del mismo tamaño, el cual representa las probabilidades de pertenencia a cada una de las clases [47].

La siguiente Figura 3.13 muestra un ejemplo de red neuronal para una tarea de clasificación en donde se utilizan capas convolucionales para la extracción de características de la imagen, una capa de atención para reducir la dimensión de la entrada y una serie de capas densas para la clasificación.

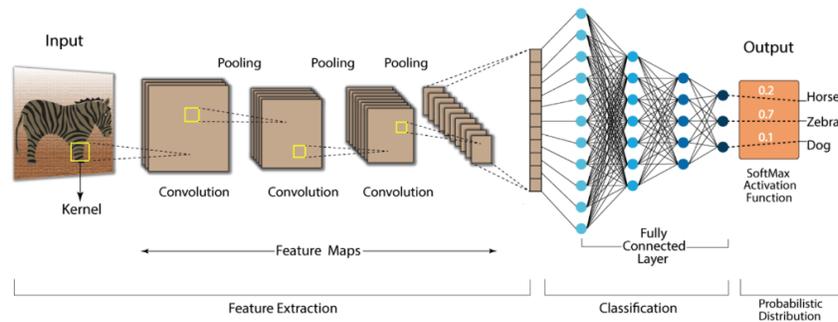


Figura 3.13: Esquematización de una red neuronal convolucional [48]

3.4.2. Proceso de aprendizaje

El proceso de aprendizaje de un modelo de *deep learning* implica la adaptación de su comportamiento a través de la modificación de las ponderaciones de cada neurona. Durante el proceso de entrenamiento, que utiliza una gran cantidad de datos con valores de salida conocidos, se parte de una ponderación arbitraria para cada neurona. Se compara el valor obtenido con el valor esperado, y en caso de discrepancia, se ajustan las ponderaciones para mejorar la concordancia. Estos ajustes se realizan mediante una función de optimización, que busca que las predicciones se asemejen cada vez más a los valores reales, facilitando así el proceso de aprendizaje. Los pasos detallados a continuación describen cómo lograr un aprendizaje satisfactorio del algoritmo mediante este procedimiento.

Inicialización de pesos

Cada neurona posee pesos que aplican ponderaciones a las conexiones, destacando las más importantes para una predicción precisa. Estos pesos pueden ser inicializados de diversas maneras, siendo una de ellas la inicialización aleatoria. Esta estrategia contribuye a que el algoritmo converja sin quedar atrapado en óptimos locales [49].

Propagación hacia adelante

Después de asignar los pesos iniciales a las neuronas, se introduce en la red la sección del conjunto de datos destinada al entrenamiento. La información atraviesa las diversas capas y se propaga en función de los pesos asignados. En la salida de la última capa se obtiene una predicción [50].

Función de pérdida

Para poder mejorar el modelo entre iteraciones de aprendizaje, es necesario contar con una función de pérdida o costo. Ésta calcula el error, es decir, la diferencia entre el valor obtenido con el valor real. Al minimizar esta función, se logra mejorar el desempeño de la red. En problemas de clasificación, se suele usar la función de pérdida de entropía cruzada, que puede ser utilizada tanto para problemas binarios como multiclase. Esta función evalúa y penaliza casos con respecto a su probabilidad de estar en la clase incorrecta [51]. Otra función de pérdida conocida es la SVM (*support vector machines*) binaria, en el cual se considera un margen entre clases, y utiliza una pérdida bisagra. Ésta penaliza a los casos incorrectos con respecto a cuán lejos se encuentren de dicho margen, mientras que los casos correctos tienen una penalización nula [52]. Una variante de esta función es la SVM binaria suavizada, donde se utiliza una versión suavizada de la pérdida bisagra. En este caso, también se aplica una penalización leve a las predicciones correctas, la cual disminuye al alejarse del margen. Esta variante hace que la función sea menos sensible al ruido de los datos [53].

Retropropagación del error

Una vez calculada la función de pérdida, es necesario ajustar los pesos de la red para que, en una nueva iteración, la función de pérdida disminuya. Se recorre la red, con el fin de calcular el gradiente de la función de pérdida con respecto a cada uno de los parámetros del modelo. Esto deja en evidencia qué parámetros son los responsables del desempeño del mismo. El ajuste de los parámetros es llevado a cabo por un optimizador que utiliza el gradiente calculado para actualizar los parámetros en cada iteración. La magnitud de la actualización está dada por la tasa de aprendizaje, o *learning rate*. Este valor tiene que ser lo suficientemente grande como para que el algoritmo no quede en un mínimo local ni sea demasiado lento, y lo suficientemente chico como para que converja [50].

El optimizador Adam combina elementos de varios optimizadores populares y se distingue por su capacidad para emplear tasas de aprendizaje adaptativas para cada parámetro individual. Utiliza dos momentos exponenciales, uno para el promedio de los gradientes y otro para el promedio de los cuadrados de los gradientes, los cuales se actualizan en cada iteración para ajustar dinámicamente la tasa de aprendizaje de cada parámetro. La inclusión de sesgo de corrección contrarresta el sesgo hacia cero

al inicio del entrenamiento. En síntesis, Adam destaca por su eficiencia y capacidad para adaptarse de manera personalizada a los parámetros, lo que lo convierte en una opción versátil para diversos problemas de optimización en el aprendizaje automático [54].

La cantidad de iteraciones, también conocidas como épocas (*epochs*), requeridas durante el entrenamiento varía según el modelo. Algunos autores optan por un número fijo de iteraciones, mientras que otros eligen un rango y evalúan la función de pérdida para decidir si se deben realizar más iteraciones. En el contexto de la validación cruzada, se evalúa la función de pérdida en un conjunto de validación desconocido para el modelo, y se selecciona la iteración en la que esta función alcanza su mínimo.

3.4.3. Distintos tipos de entrenamiento

Dentro del rubro de la IA, existe una clasificación sobre los niveles de supervisión que requiere el algoritmo para su entrenamiento. Los algoritmos supervisados cuentan con etiquetas para cada una de sus datos. El algoritmo aprende al observar los datos y sus etiquetas, para luego predecir la etiqueta de un nuevo dato. Por otro lado, los algoritmos no supervisados carecen de etiquetas y buscan patrones internos para realizar la clasificación. Por ejemplo, un algoritmo no supervisado entrenado con imágenes de radiografías podría utilizar parámetros como tamaño y forma para distinguir diferentes partes del cuerpo. Mientras que los algoritmos supervisados requieren etiquetas proporcionadas por expertos, los no supervisados suelen demandar una mayor cantidad de datos para su funcionamiento, lo que los hace computacionalmente más costosos [55]. A continuación, en la Figura 3.14, se puede observar un esquema de esta clasificación y en qué tipo de tareas se utiliza cada una.

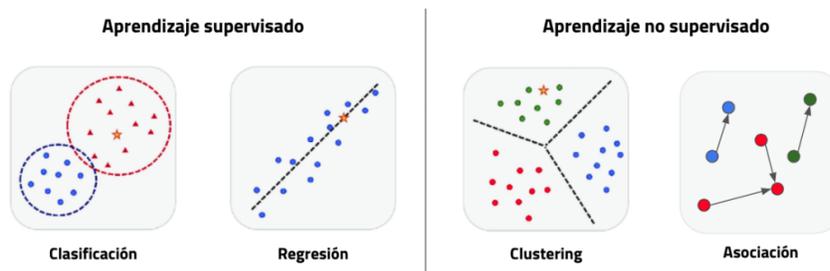


Figura 3.14: Tareas de aprendizaje supervisado y no supervisado

Existe una categoría intermedia denominada algoritmos débilmente supervisados. Estos algoritmos se basan en el uso de etiquetas débiles o genéricas aplicadas a todos los datos, proporcionando clasificaciones sin detalles específicos. Aunque estas etiquetas tienden a tener más ruido y carecen de gran especificidad, simplifican considerablemente el proceso de etiquetado. Este tipo de supervisión permite llevar a cabo tareas de clasificación que requieren un análisis más detallado por parte del algoritmo, llevándolo a aprender qué aspectos específicos debe considerar para realizar la clasificación [56].

Una forma de aprendizaje débilmente supervisado es el aprendizaje de múltiples instancias (MIL por sus siglas en inglés *Multiple Instance Learning*). En este tipo de modelo, las instancias de aprendizaje se encuentran organizadas en grupos llamados bolsas. En lugar de tener una etiqueta para cada instancia, se tiene una etiqueta para cada bolsa. En este modelo, se trabaja bajo la asunción de que en una bolsa negativa, todas las instancias son negativas, mientras que en una bolsa positiva, hay al menos una instancia positiva [57]. En el caso de estar trabajando con imágenes, los píxeles o parches son instancias mientras que la imagen es la bolsa, y se tiene una etiqueta para la imagen entera. Este tipo de modelo es muy útil para problemas de clasificación de imágenes médicas, donde la presencia de un evento biológico en algún sector de la imagen hace que la imagen entera sea positiva. Se puede observar un ejemplo de esto en la Figura 3.15 a continuación.

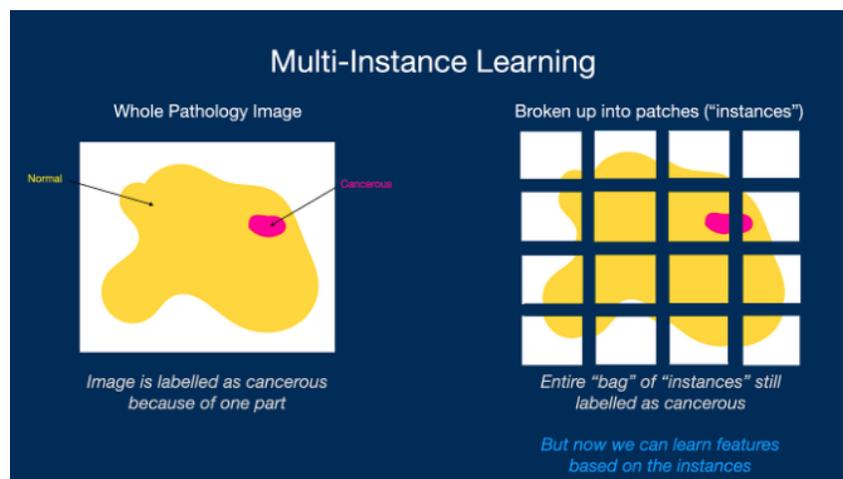


Figura 3.15: Esquemática de aprendizaje de múltiples instancias [58]

3.4.4. Aprendizaje por transferencia

En el contexto del aprendizaje por transferencia, el proceso implica aprovechar el conocimiento adquirido por un modelo durante su entrenamiento en una tarea fuente (*source task*) para mejorar su rendimiento en una tarea destino (*target task*). Durante la fase de pre-entrenamiento en la tarea fuente, el modelo aprende a extraer características relevantes y a entender patrones complejos presentes en los datos. Luego, en la fase de ajuste o *fine-tuning* en la tarea destino, el modelo utiliza este conocimiento previo como punto de partida.

En el ajuste fino, se adapta el modelo pre-entrenado a la tarea destino mediante la actualización de sus pesos a medida que se expone al nuevo conjunto de datos. Es común "descongelar" algunas capas del modelo para permitir que se ajusten más específicamente a los detalles de la nueva tarea, mientras se mantienen otras capas con los pesos pre-entrenados para retener el conocimiento general aprendido. Este enfoque es particularmente útil cuando el conjunto de datos de la tarea destino es limitado, ya que el modelo aprovecha las características aprendidas en la tarea fuente para mejorar su desempeño en la nueva tarea.

El aprendizaje por transferencia permite una mayor eficiencia y rendimiento al evitar la necesidad de entrenar modelos desde cero en cada tarea, especialmente cuando se enfrenta a limitaciones de datos o recursos [59].

ResNet-50

Un modelo frecuentemente utilizado para el aprendizaje por transferencia es ResNet-50. ResNet significa Red Residual y es un tipo específico de CNN introducido en el artículo "*Deep Residual Learning for Image Recognition*" publicado en 2015 por He Kaiming et al. [60]. Esta red neuronal está compuesta por 50 capas: 48 capas convolucionales, una capa de *max-pooling* y una de *average-pooling*. Como se mencionó anteriormente, las capas convolucionales son esenciales en la extracción de características de una imagen mientras que las dos últimas capas se utilizan para reducir y resumir las características de las entradas. La característica distintiva de las redes neuronales residuales es el uso de bloques residuales, que permiten que las capas de una red aprendan las diferencias entre las características originales y las características transformadas. Esto facilita el entrenamiento de redes muy profundas como ResNet-50 [61].

Vision Transformers

Los *Vision Transformers* (ViT), también conocidos como Transformadores de Visión, representan un tipo de modelo basado en redes neuronales diseñado especialmente para abordar tareas relacionadas con el procesamiento de imágenes. Inicialmente concebidos como una arquitectura de aprendizaje profundo para el procesamiento del lenguaje natural, estos modelos fueron posteriormente adaptados para sustituir a las CNN en las tareas de procesamiento de imágenes por los investigadores Dosovitskiy et al. en el año 2020 [62].

Los transformadores tradicionales emplean mecanismos de atención durante el proceso de entrenamiento del modelo. En la arquitectura ViT, el primer paso consiste en dividir la imagen en parches, comúnmente de dimensiones 16x16 píxeles. Cada parche se transforma en un vector (*embedding*) a través de una CNN, capturando así las características relevantes. La elección de una CNN se justifica por su capacidad para extraer características significativas de los parches. Estos vectores se introducen luego en un codificador transformador, que puede conceptualizarse como una serie de capas de atención en donde el modelo busca las relaciones entre los parches y aprende cómo los mismos contribuyen al etiquetado final de la imagen [63].

De esta manera, la salida del codificador transformador son vectores con las características de la imagen. Éstos son decodificados por una red de múltiples capas para poder realizar una clasificación de la imagen. Los ViT tienen la ventaja de poder procesar parches, lo cual es menos costoso computacionalmente que procesar pixel por pixel, pero sin perder de vista dónde se ubican éstos y cómo se vinculan entre sí. Los ViT pueden ser entrenados con conjuntos de imágenes más pequeños y se pueden utilizar para diversas funciones de procesamiento de imágenes, desde clasificación hasta detección de objetos. En la Figura 3.16 se puede observar una esquematización de un modelo ViT [63].

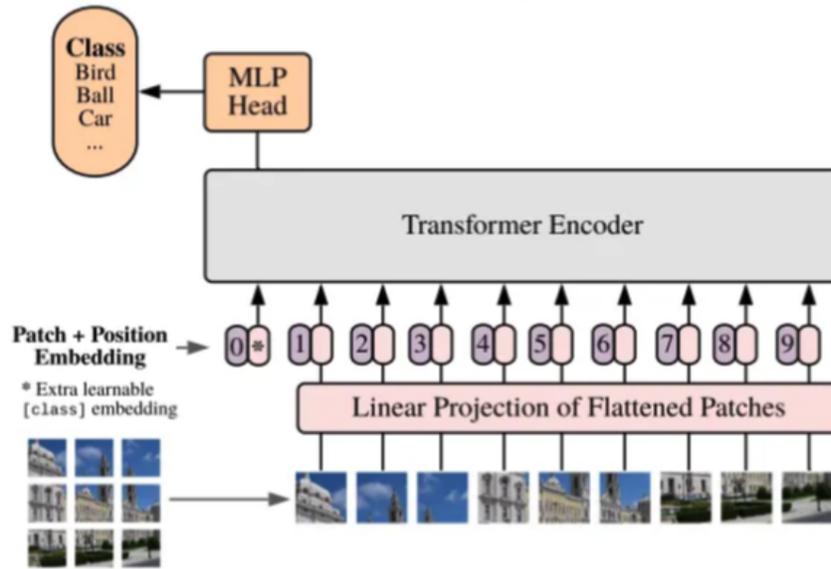


Figura 3.16: Esquematzación de un ViT [63]

3.5. Evaluación y métricas

Una vez que el algoritmo ha sido entrenado, se procede a realizar pruebas para evaluar su desempeño. Este proceso se lleva a cabo utilizando un conjunto de datos de evaluación, para el cual se conocen los resultados reales. Se compara la salida del modelo con estos resultados para obtener métricas estadísticas que proporcionan información sobre su rendimiento. Como se ilustra en la Figura 3.17, en el caso de tareas binarias, los resultados de la evaluación se clasifican en cuatro categorías según la combinación de los valores reales y las predicciones. Esta representación se conoce como una matriz de confusión.

		Predicciones	
		Negativo	Positivo
Valores Reales	Negativo	Verdadero Negativo (VN)	Falso Positivo (FP)
	Positivo	Falso Negativo (FN)	Verdadero Positivo (VP)

Figura 3.17: Matriz de confusión

La exactitud o *accuracy* es la tasa de casos predichos correctamente sobre el total de casos. Es la más simple de todas las métricas, y habla del desempeño global del

algoritmo. La fórmula para esta métrica se encuentra en la ecuación (3.1).

$$Exactitud = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.1)$$

La sensibilidad o *recall* es la métrica que mide la utilidad de la prueba al ser utilizada en casos positivos. Es la tasa de verdaderos positivos con respecto al total de los casos positivos. En cambio, la especificidad mide la utilidad de la prueba al ser utilizada en casos negativos. Es la tasa de verdaderos negativos con respecto al total de los casos negativos. Las ecuaciones que describen a estas métricas se pueden observar en las ecuaciones (3.2) y (3.3) respectivamente.

$$Sensibilidad = \frac{VP}{VP + FN} \quad (3.2)$$

$$Especificidad = \frac{VN}{VN + FP} \quad (3.3)$$

Luego, existe la precisión o valor predictivo positivo (PPV). La misma es la tasa de verdaderos positivos con respecto al total de los casos predichos como positivos. Por otro lado, el valor predictivo negativo (NPV) es la tasa de verdaderos negativos con respecto al total de los casos predichos como negativos. Las ecuaciones que describen a estas métricas se pueden observar en las ecuaciones (3.4) y (3.5) respectivamente.

$$PPV = \frac{VP}{VP + FP} \quad (3.4)$$

$$NPV = \frac{VN}{VN + FN} \quad (3.5)$$

Por último, estas métricas se pueden contrastar para obtener más información. El análisis ROC (característica operativa del receptor) es una gráfica entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (1-especificidad) a distintos umbrales de decisión. Calculando el área debajo de esta curva, se obtiene la métrica AUROC (área debajo de la curva ROC). Esta métrica aporta información sobre la habilidad del algoritmo de distinguir entre las clases. Un valor de AUROC igual a 1 indica un modelo perfecto, mientras que un valor de 0,5 sugiere que el modelo es tan bueno como una clasificación aleatoria [64]. Similarmente, existe la AUPR (área debajo de la curva PR), una métrica derivada de calcular el área debajo de la curva PR (precisión-recall). Esta curva grafica el valor de la sensibilidad y la

precisión para los distintos umbrales de decisión [65]. Por lo tanto, la curva PR y la AUPR son muy útiles para encontrar un umbral donde tanto la precisión como la sensibilidad son satisfactorias. En la Figura 3.18 se pueden observar ambos tipos de curvas.

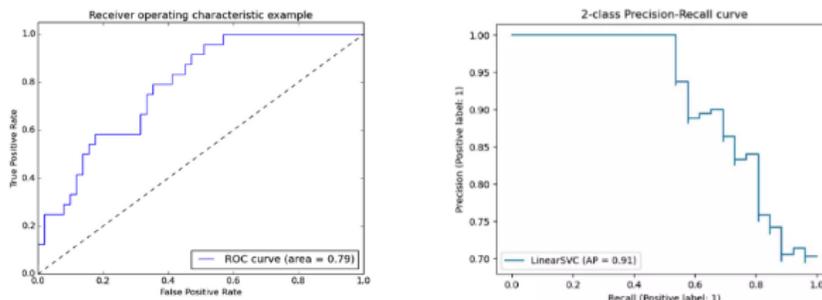


Figura 3.18: Curva ROC y curva PR [66]

La decisión sobre qué métricas considerar a la hora de evaluar un algoritmo depende del tipo de tarea que se está realizando y de su base de datos. En los casos clínicos, es importante considerar qué es más peligroso para un paciente, si un falso positivo o un falso negativo. Además, en el caso de contar con clases desbalanceadas, es necesario utilizar las métricas que tengan esto en consideración.

3.6. Patología digital

La patología digital es un entorno dinámico basado en imágenes, que permite la adquisición, gestión e interpretación de información patológica generada a partir de un portaobjetos digitalizado [1]. El término para las imágenes en el ordenador es *Whole Slide images* o diapositivas completas de preparado histológico (WSI por sus siglas en inglés). El proceso de digitalización de un corte histológico implica el uso de un escáner especializado en histopatología. Este escáner realiza un barrido minucioso de la lámina histológica, capturando imágenes de alta resolución en múltiples secciones. Estas imágenes se ensamblan digitalmente para crear un WSI completo. Durante este proceso, se registran detalles microscópicos, y la información resultante se almacena en formato digital para su visualización y análisis en una pantalla de computadora [67]. En la Figura 3.19 se observan tres ejemplos de este tipo de imágenes, WSI de cáncer endometrial.

Estas imágenes digitales, obtenidas a partir de biopsias o resecciones tumorales,

pueden luego ser almacenadas, compartidas y analizadas digitalmente, ofreciendo una serie de ventajas sobre los métodos tradicionales de patología. Se ofrece la posibilidad de acceso desde cualquier sitio, la colaboración entre patólogos en tiempo real y la utilización de herramientas digitales de análisis de imágenes para mejorar la precisión de diagnóstico [68].

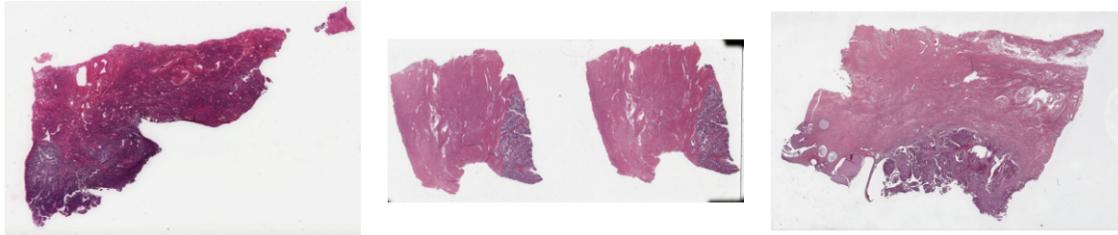


Figura 3.19: WSI de cáncer de endometrio

Los enfoques de patología computacional basados en aprendizaje profundo requieren, o bien la anotación manual del preparado en su totalidad, en un entorno totalmente supervisado, o grandes cantidades de datos con etiqueta a nivel de la diapositiva en un entorno débilmente supervisado. Las imágenes H&E se suelen digitalizar en diferentes resoluciones y luego se guardan como una única imagen en un archivo SVS (*Slide Viewer System*). Dichos archivos contienen imágenes piramidales con distintos niveles de magnificación según la digitalización. Cada nivel corresponde con una resolución, es decir cantidad de micrones por pixel (mpp), que indica la magnificación máxima a la cual se puede observar la imagen sin perder calidad. En la mayoría de los casos, el nivel más bajo corresponde a la mayor magnificación mientras que el más alto, a la menor. Este formato de imagen de alta resolución tiene características muy distintas a una imagen en formato JPG o PNG, por lo que su procesamiento es muy distinto.

La manipulación integral de imágenes piramidales en formato SVS a menudo resulta desafiante, por lo que es común segmentar algún nivel de magnificación según las necesidades y convertirlo en parches de dimensiones constantes. Optar por una magnificación más alta permite examinar con mayor detalle estructuras diminutas, mientras que una magnificación menor abarca una mayor extensión de tejido, y reduce el costo computacional asociado al procesamiento de los parches.

En la actualidad, existen numerosos enfoques para realizar operaciones sobre estas imágenes, pero se busca comúnmente métodos que no requieran la extracción

manual de regiones de interés (ROI) ni anotaciones detalladas a nivel de píxeles o parches. El objetivo es obtener resultados sólidos incluso con una cantidad limitada de muestras. Dado que no siempre se encuentra accesible la información genómica, transcriptómica o proteómica de las muestras, resulta crucial hallar modelos que gestionen eficientemente conjuntos de datos y puedan operar con una cantidad reducida de imágenes.

3.7. Arquitectura CLAM

La herramienta CLAM (*clustering-constrained-attention multiple instance learning*) fue desarrollada por Mahmood Lab y publicada en marzo de 2021 como una herramienta para resolver los principales problemas de la patología computacional mencionados previamente [69]. Dicho laboratorio pertenece a la institución *Brigham and Women's Hospital* (BWH) en Boston, Massachusetts, Estados Unidos y se dedica a la utilización de inteligencia artificial para el análisis de imágenes patológicas [70].

CLAM es una herramienta de aprendizaje profundo de alto rendimiento para la clasificación débilmente supervisada en patología computacional. En este enfoque, cada diapositiva digitalizada de tejido patológico se etiqueta con un diagnóstico, pero no se proporciona ninguna anotación a nivel de píxel o región. Esto significa que CLAM no necesita extraer ROI ni realizar anotaciones de parches. Mediante tres análisis independientes, los autores demostraron que la herramienta puede obtener un alto rendimiento a través de diferentes tareas de clasificación. Además ha demostrado ser eficiente en cuanto al manejo de datos ya que dichos resultados prometedores se mantienen aún reduciendo el número de datos etiquetados. La herramienta se evalúa en tres tareas independientes: clasificación de carcinoma de célula renal (RCC), clasificación de cáncer de pulmón de células no pequeñas (NSCLC) y la detección de tumor en nódulos linfáticos. Se utilizan tanto datasets públicos como cohortes independientes y se ha demostrado que, a pesar de haber sido entrenada con resecciones tumorales, la misma puede ser adaptada para biopsias e incluso imágenes fotomicrográficas obtenidas con un celular inteligente. Las imágenes fueron obtenidas de *The Cancer Genome Atlas* (TCGA). A continuación se puede observar la Tabla 3.1 en la cual se resumen los resultados obtenidos en cada una de estas pruebas. La métrica provista corresponde al promedio del AUROC

junto a su desvío estándar de los 10 pliegues sobre los cuales se realizó la evaluación.

Tarea	Dataset	Cantidad de imágenes	Magnificación	AUROC
Clasificación RCC	TCGA [71]	884	x20	0.991±0.004
	interno	135	x20	0.972±0.008
Clasificación NSCLC	TCGA [72]	1967	x20	0.956±0.020
	interno	24	x20	0.975±0.007
Detección de metástasis	CAMELYON [73]	899	x40	0.953±0.029
	interno	133	x40	0.940±0.015

Tabla 3.1: Desempeño de arquitectura CLAM en distintas tareas

Desarrollo

El objetivo de este trabajo es desarrollar una herramienta de patología digital para la clasificación de carcinomas endometriales proficientes y deficientes para el sistema de reparación del ADN. Con esto en mente se busca desarrollar un modelo de aprendizaje profundo que pueda cumplir con dicha tarea. En este trabajo se implementan y evalúan cuatro modelos distintos y luego se elige el que mejor se adapta a los requisitos de esta clasificación. A grandes rasgos, el desarrollo y evaluación de un modelo consiste en entrenarlo con un conjunto de imágenes de las cuales se conoce su estado de inestabilidad microsatelital, es decir, se sabe si cada una de ellas es MSI o MSS. La idea es que el modelo aprenda cuáles son las características distintivas de cada clase y que cuando se le presente una imagen con estado desconocido, el mismo pueda predecir su estabilidad o inestabilidad microsatelital. Los cuatro modelos comparten la misma arquitectura y se diferencian únicamente en el conjunto de datos que se utiliza para su entrenamiento. Las diferencias radican en la magnificación de las imágenes y en qué tan balanceadas están ambas clases en el conjunto de entrenamiento. Los cuatro modelos son los siguientes:

	Magnificación	Equilibrio de clases
modelo_balanceado_20x	20x	balanceado
modelo_desbalanceado_20x	20x	desbalanceado
modelo_balanceado_10x	10x	balanceado
modelo_desbalanceado_10x	10x	desbalanceado

Tabla 4.1: Distintos modelos de entrenamiento

En primer lugar, es importante destacar que los cuatro modelos se basan en la arquitectura CLAM que no sólo provee un modelo de aprendizaje profundo para la clasificación sino que también facilita cada uno de los pasos en el procesamiento

de las imágenes patológicas. El flujo de trabajo completo para el desarrollo de los modelos puede observarse en la Figura 4.1 y se explica a continuación.

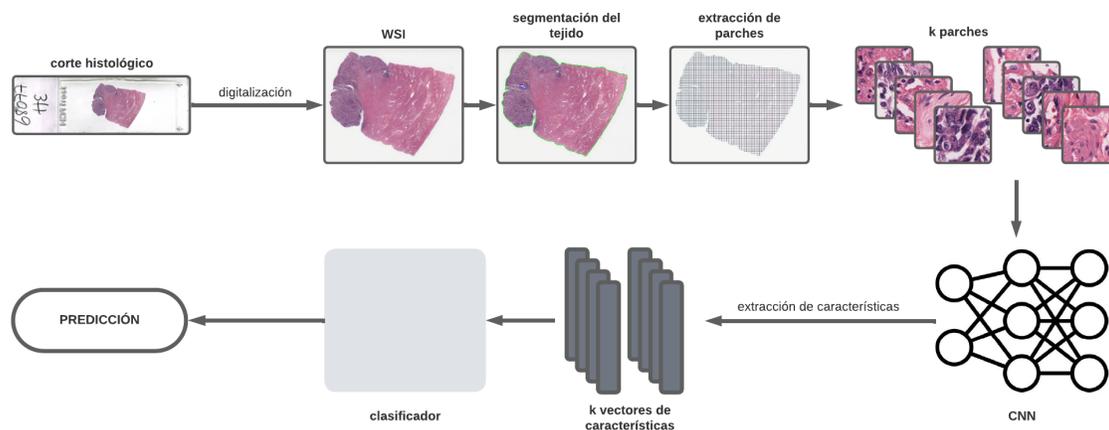


Figura 4.1: Flujo de trabajo de la herramienta completa

El primer paso, tanto para entrenar al modelo como para evaluar el mismo en una imagen nueva, es segmentar el tejido dentro de la imagen. Luego, se segmenta dicha región en parches iguales de 256x256 píxeles para poder operar sobre regiones más pequeñas de la misma.

Estos parches ingresan a una CNN pre entrenada para poder extraer 1024 características de cada uno. Son estas características las que van a ingresar al modelo de clasificación para poder predecir la estabilidad o inestabilidad microsatelital para cada WSI. Al trabajar en esta dimensión de características en vez de en la dimensión de los píxeles, se reduce significativamente el costo computacional.

En cuanto al modelo de clasificación, para poder entrenar un modelo de aprendizaje profundo utilizando WSI sin anotaciones ni extracción de ROIs, CLAM utiliza una función de agregación basada en atención para combinar las características de los parches de una imagen en una representación a nivel de toda la diapositiva. Tanto durante el entrenamiento como en la evaluación, el modelo examina y ordena todos los parches de tejido asignándoles un *score*/puntaje de atención a cada uno de ellos. Este puntaje provee información sobre la contribución o importancia de dicho parche en la representación de toda la imagen para cada clase. La salida del modelo es la probabilidad que presenta la diapositiva completa de pertenecer a cada una de las clases junto con su clase predicha.

4.1. Elección de la arquitectura CLAM como base para el modelo

La selección de CLAM como arquitectura principal se basó en una búsqueda y evaluación exhaustiva de diversas alternativas. Esta opción se destacó por su prometedor rendimiento, eficiencia computacional, manejo eficaz de datos y la mínima intervención de patólogos profesionales. A continuación, se detallan las ventajas que ofrece esta arquitectura en comparación con otras opciones consideradas.

En primer lugar, la plataforma fue creada para procesar WSI de muestras patológicas con tinción H&E en formato de imagen piramidal de alta resolución. La manipulación de estas imágenes es sumamente compleja y, por lo tanto, es importante contar con un algoritmo que emplee las librerías adecuadas para poder procesar este tipo de archivos y sacar provecho de la alta resolución del formato piramidal. El procesamiento de una imagen patológica requiere de varios pasos y CLAM facilita cada uno de ellos.

En segundo lugar, al tratarse de una metodología semi-supervisada, CLAM facilita la obtención de muestras. Sólo es necesaria una etiqueta para cada imagen, indicando si ésta es estable o inestable. En el caso de utilizar un algoritmo supervisado, sería necesario que un patólogo experto indique qué sector de la imagen es relevante para la tarea, ya sea con un ROI o etiquetando cada uno de los parches. Ya que la imagen incluye tanto tejido tumoral como tejido sano, y considerando que la región tumoral es la que contiene información relevante, la tarea de preparar los casos para el entrenamiento se simplifica notablemente cuando se requiere únicamente una etiqueta para cada imagen.

En tercer lugar, la forma en la que se procesan los datos está diseñada con el objetivo de reducir el costo computacional. Como fue mencionado, las imágenes piramidales son de muy alta resolución, lo cual las hace muy grandes, por lo que procesarlas suele ser muy costoso computacionalmente. En esta arquitectura se plantea un paso previo al entrenamiento, en el cual se extraen características importantes de cada parche y se transforman en un espacio numérico con el cual es más sencillo trabajar. Esto logra que el entrenamiento sea mucho más rápido y computacionalmente menos costoso.

Por último, CLAM no se trata de un modelo en el cual todos los parámetros fueron determinados por los autores, sino que es posible intervenir significativamente

en cada paso del proceso de aprendizaje y evaluación. De esta manera es posible adaptar el modelo a distintas tareas con distintas necesidades. Es posible supervisar el proceso en cada momento, para poder corregirlo y lograr que el algoritmo obtenga buenos resultados.

Es fundamental resaltar que este estudio representa un desafío significativo, ya que la tarea de clasificación resulta considerablemente más compleja que aquellas para las que CLAM fue originalmente concebido. Esta herramienta fue desarrollada y evaluada para clasificar tumores a nivel de microscopía de luz, según características morfológicas visibles por un médico experto en el área. Nuestro trabajo, altamente ambicioso, busca la detección de fenómenos moleculares, tan sólo detectables por técnicas complementarias sofisticadas de laboratorio.

A continuación se detalla el proceso de implementación de esta arquitectura, incluyendo la obtención de muestras para su entrenamiento y evaluación. Además, se desarrolla un prototipo de interfaz gráfica para que esta herramienta pueda ser implementada y operada por los patólogos en una institución médica.

4.2. Obtención de muestras

Para llevar a cabo este trabajo, se ensamblaron dos conjuntos de imágenes histológicas de carcinomas endometriales de las cuales se conoce su estado microsatelital. Por un lado, para el entrenamiento y una primera evaluación del modelo, se creó el conjunto con muestras del TCGA y del *Clinical Proteomic Tumor Analysis Consortium* (CPTAC). Este conjunto es denominado TCGA-CPTAC, y cuenta con muestras de los proyectos TCGA-UCEC y CPTAC-UCEC. Las siglas UCEC corresponden al proyecto de carcinoma endometrial del cuerpo uterino (*Uterine Corpus Endometrial Carcinoma*). Por otro lado, para contar con una instancia de evaluación independiente, se creó un segundo conjunto con imágenes obtenidas de muestras retrospectivas de un laboratorio independiente de anatomía patológica en Buenos Aires.

4.2.1. Conjunto TCGA-CPTAC

Tanto el TCGA como el CPTAC son dos proyectos importantes en la investigación del cáncer, llevados a cabo por el Instituto Nacional del Cáncer de Estados Unidos. El TCGA está enfocado en las caracterizaciones genéticas y epigenéticas del cáncer,

mientras que el CPTAC se enfoca en la proteómica del mismo. Por más que su enfoque sea distinto, ambas tienen como objetivo la recopilación de información asociada al cáncer, y hacerla disponible al público para una mejor comprensión y un avance en la investigación del área.

Estas bases de datos fueron seleccionadas como fuentes de imágenes por su presencia recurrente en otros trabajos similares. La investigación de Hong et al. [74] utiliza tanto TCGA-UCEC como CPTAC-UCEC para el entrenamiento de su modelo de aprendizaje profundo destinado a la clasificación del cáncer de endometrio en sus distintos subtipos moleculares. Fremont et al. [75] emplea diversas fuentes de datos para abordar la misma tarea de clasificación de carcinomas endometriales, siendo TCGA-UCEC el único conjunto de datos de acceso público en dicha investigación. También es importante considerar que CLAM, si bien se enfoca en otros órganos, también fue entrenado y evaluado con muestras del TCGA. Éstos son unos de los únicos conjuntos de datos abiertos al público, ya que si bien muchos hospitales o centros médicos cuentan con imágenes digitalizadas y con su estado microsatelital, esta información no es de público acceso.

Para cada uno de los conjuntos de datos, la forma en la que se evalúa la inestabilidad microsatelital es distinta. El TCGA, que tiene un enfoque genómico, utiliza técnicas de biología molecular para determinar si el microsatélite es estable o inestable. Se emplean siete locus, los cinco recomendados por el NCI, junto con BAT40 y TGFBR2. En este caso, para que la muestra sea clasificada como MSI-H, se requiere que tres o más marcadores presenten inestabilidad. Si sólo uno o dos resultan inestables, la muestra se clasifica como MSI-L, y si todos resultan estables, se designa como MSS [76]. En los estudios previos realizados por Hong et al. y Fremont et al. las muestras MSI-L son agrupadas junto a las MSS ya que no se consideran a las mismas como tumores inestables. En este trabajo, para contribuir con el balance de las clases y evitar generar confusión en el modelo, se excluyen los casos MSI-L. Es importante destacar que los mismos representan menos del 4% del total de los casos.

En cambio, el CPTAC, que tiene un enfoque proteómico, evalúa el mecanismo MMR para determinar el estado microsatelital. Si el mismo resulta deficiente, se cataloga a la muestra como MSI mientras que si resulta proficiente, como MSS. La evaluación de dicho mecanismo implica el análisis de las cuatro proteínas responsables. En el caso de que alguna de ellas esté ausente, la muestra se categoriza como

MSI-H [77].

En ambos proyectos, TCGA-UCEC y CPTAC-UCEC, las imágenes fueron digitalizadas con un digitalizador Aperio. Lamentablemente, aunque las imágenes sean de público acceso, no toda la información genómica y proteómica se encuentra habilitada. Esto limita significativamente la cantidad de muestras utilizadas.

Todas las imágenes fueron analizadas por una patóloga especializada para poder aplicar un criterio de exclusión sobre las muestras y eliminar aquellas que no se encontraban en condiciones aptas para su incorporación en el algoritmo. Se consideraron tres factores para este criterio: que la muestra no contenga grandes tachaduras con marcador indeleble, que no posea bálsamo de Canadá que dificulte la visualización correcta del tejido en cuestión y que no contenga pliegues en el corte histológico que impida la lectura. A continuación en la Figura 4.2 se puede observar un ejemplo de cada uno de estos casos.

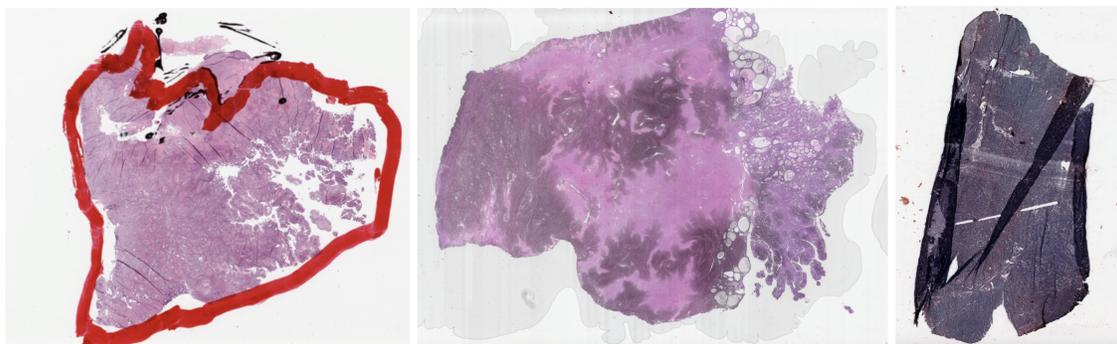


Figura 4.2: Ejemplos de imágenes con marcador, bálsamo y grandes pliegues

Una vez aplicados los criterios de exclusión a las muestras del TCGA, se obtuvo un conjunto de 345 imágenes pertenecientes a 320 pacientes. 44 de estas imágenes tienen una magnificación nativa de 20x y una resolución de 0,5 micrones por pixel (mpp), mientras que las 301 imágenes restantes tienen una magnificación nativa de 40x y una resolución de 0,25 mpp. Del total de las imágenes del TCGA, 124 son clasificadas como MSI-H y 221 como MSS.

Luego, se aplicaron los mismos criterios de exclusión a las muestras del CPTAC para obtener un conjunto de 159 imágenes de 82 pacientes. Todas estas imágenes tienen una magnificación nativa de 20x y una resolución de 0,5 mpp. 38 de estas imágenes son MSI-H y 121 MSS.

Al unir las muestras del TCGA y CPTAC, se creó un único conjunto TCGA-

CPTAC con 504 imágenes, de las cuales 162 son MSI-H y 342 MSS. Es importante destacar que para entrenar a un modelo, todas las imágenes deben tener la misma magnificación. Por lo tanto, se utilizó una magnificación máxima de 20x para entrenar y evaluar los distintos modelos. La formación de un conjunto de entrenamiento derivado de dos fuentes independientes, las cuales recopilan casos de diversas instituciones, confiere la ventaja de introducir variabilidad en dicho conjunto. Este enfoque se implementa con el propósito de mitigar el riesgo de sobreajuste del modelo a las imágenes de entrenamiento, permitiendo, de este modo, que el modelo sea capaz de discernir características en imágenes, independientemente de su procedencia. Para poder validar y evaluar el proceso de aprendizaje, el conjunto de imágenes fue dividido en 3 subconjuntos. El 80 % de las imágenes fue incluido en el conjunto de entrenamiento, el 10 % en el conjunto de validación y el 10 % en el conjunto de evaluación. Para cada subconjunto se mantuvo aproximadamente la misma proporción de muestras estables e inestables que en el conjunto total.

Como se puede notar, las clases para este conjunto de datos se encuentran muy desbalanceadas, contando con más del doble de imágenes MSS que MSI-H. Por lo tanto, se busca crear una versión balanceada del mismo para entrenar ambos modelos en paralelo y determinar cuál da mejores resultados. El conjunto TCGA-CPTAC balanceado cuenta con 382 imágenes de 314 pacientes, de las cuales 162 corresponden a tumores MSI-H y 220 a tumores MSS. Aunque la cantidad de imágenes por clase no es exactamente la misma, las cifras son más equilibradas. Si bien lo óptimo sería contar con clases totalmente balanceadas, también se persigue disponer de un conjunto lo bastante extenso como para que el modelo pueda aprender de una base de datos amplia. La Tabla 4.2 muestra la cantidad de casos en cada conjunto y subconjunto.

		Entrenamiento	Validación	Evaluación	Total
Modelo TCGA-CPTAC desbalanceado	MSI-H	130	16	16	162
	MSS	274	34	34	342
	Total	404	50	50	504
Modelo TCGA-CPTAC balanceado	MSI-H	130	16	16	162
	MSS	176	22	22	220
	Total	306	38	38	382

Tabla 4.2: Subconjuntos de entrenamiento, validación y evaluación para los modelos entrenados con TCGA-CPTAC (balanceado y desbalanceado)

4.2.2. Imágenes de archivo de un laboratorio de patología

Además, para verificar la consistencia de los resultados obtenidos al entrenar modelos con TCGA-CPTAC, se generó un conjunto de datos independiente utilizando imágenes de archivo de un laboratorio de patología, que incluyen muestras recopiladas entre los años 2019 y 2023. Dado que el objetivo a futuro de este trabajo consiste en implementar una clasificación mediante este modelo en instituciones médicas, se busca obtener resultados satisfactorios en el conjunto independiente.

Se recopilaron los casos dentro del período mencionado que ya habían sido sometidos a inmunohistoquímica para evaluar el mecanismo MMR como subrogante para predecir la inestabilidad microsatelital. Se excluyeron deliberadamente los casos con resultados indeterminados en la IHQ, siendo importante señalar que estos casos tienden a ser clasificados como MSI-L en una subsiguiente prueba por biología molecular. Se tomaron sólo aquellos casos correspondientes a biopsias o resecciones de cuerpo de útero, con diagnóstico confirmado de carcinoma endometrial. En este instituto las muestras no se encontraban digitalizadas, por lo que se llevó a cabo un proceso de digitalización con un equipo Leica Aperio LV1. Este dispositivo tiene la capacidad de procesar y mostrar simultáneamente hasta 4 muestras, ofreciendo varias opciones de magnificación que van desde 2,5x hasta 63x para la digitalización [77].

En relación con los criterios de exclusión, se aplicaron los mismos que fueron utilizados para la construcción del conjunto de datos de entrenamiento, además de un control adicional para verificar la correcta digitalización de las muestras. Se excluyeron casos en los que los cortes histológicos resultaron demasiado gruesos, lo

que dificulta que el digitalizador logre el enfoque necesario para obtener imágenes nítidas. La Figura 4.3 ilustra la diferencia entre dos imágenes digitalizadas a una magnificación de 10x, una con una digitalización adecuada y otra con una digitalización incorrecta.

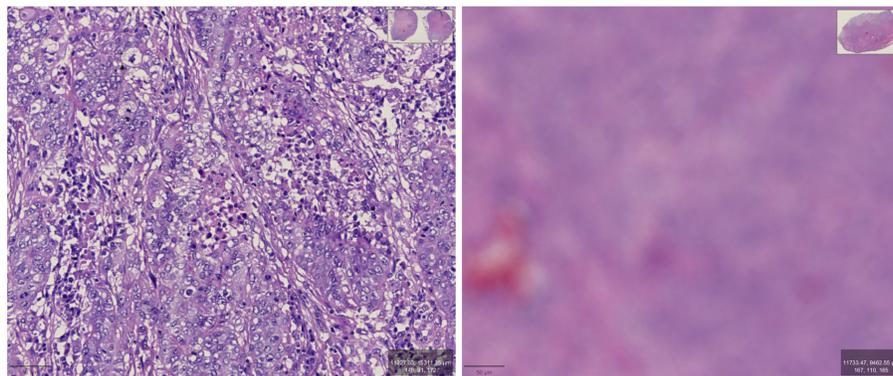


Figura 4.3: Comparación entre una buena (izquierda) y una mala (derecha) digitalización a una magnificación de 10x

Con las imágenes aprobadas se generó un conjunto de datos independiente de 51 imágenes, de las cuales 22 son MSI-H y 29 MSS. Las imágenes fueron digitalizadas a una magnificación de 20x con una resolución de 0,25 mpp. Al comparar estos valores con los del conjunto de entrenamiento, se puede notar que esta resolución corresponde a una magnificación de 40x. Lo que se busca unificar entre todas las muestras es el nivel de resolución. Por lo tanto, para evaluar ambos modelos -aquellos entrenados con imágenes en 10x y 20x- se deben redimensionar las imágenes del conjunto independiente con un factor de reducción de 4 y 2 respectivamente.

4.3. Hardware y software

Este trabajo fue llevado a cabo a través del servidor del ITBA, el cual cuenta con un CPU AMD Ryzen 7 5800X con 8 núcleos, con 32 GB de RAM y un GPU NVidia Quadro RTX 8000 con 48 GB de VRAM. En cuanto a software, toda la arquitectura CLAM es implementada en Python (versión 3.7.5) y utiliza varias librerías para llevar a cabo sus funciones. La instalación de CLAM incluye la instalación de las librerías importantes en sus versiones correspondientes, algunas de ellas siendo openslide (versión 3.4.1), opencv (versión 4.1.1), pillow (versión 6.2.1), Pythorch (versión 1.3.1), matplotlib (versión 3.1.1), seaborn (versión 0.8.1) y scikit-

learn (versión 0.22.1).

4.4. Etapas del flujo de trabajo

4.4.1. Segmentación del tejido

Para cada diapositiva digitalizada, el flujo de trabajo comienza con una segmentación automática de la región de tejido. La diapositiva se lee en la memoria a una resolución reducida por un factor de reducción de 64x y se convierte del espacio de color RGB (*Red Green Blue*) a HSV (*Hue Saturation Value*). Se calcula una máscara binaria para identificar las regiones de tejido de la imagen examinando el canal de saturación de la misma. La saturación es un componente del espacio de color HSV que mide la pureza del color. La umbralización implica establecer un valor límite, de modo que los píxeles con saturación por encima de ese límite se consideran parte del tejido, mientras que los píxeles con saturación por debajo se consideran como fondo. Antes de la umbralización la imagen se “filtra” con un filtro de mediana para suavizar los bordes y reducir el ruido. Finalmente, se realiza un cierre morfológico adicional para rellenar pequeños espacios y agujeros en la máscara binaria. El cierre morfológico es una operación que se utiliza en procesamiento de imágenes para unir regiones cercanas y eliminar pequeños vacíos.

Tanto el umbral para la saturación como el tamaño del filtro de mediana y del cierre morfológico, son parámetros que se pueden modificar para cada imagen según sus características. Debido a las diferencias en la digitalización de los distintos conjuntos de datos, para cada uno de ellos -TCGA, CPTAC y el independiente- se utilizaron los parámetros que mejor segmentan el tejido. Dichos parámetros se encuentran en la Tabla 9.1 del anexo. Los autores de CLAM establecen el valor de los parámetros óptimos para imágenes del TCGA. La diferencia entre el TCGA y el CPTAC radica principalmente en que la mayoría de las imágenes del TCGA son resecciones mientras que las del CPTAC son biopsias. En general, una biopsia es una muestra tumoral mucho más pequeña que una resección en donde se suele obtener una región mucho más grande de tejido [79]. En las siguientes imágenes se puede observar la diferencia entre una biopsia y una resección.



Figura 4.4: Ejemplo de una resección

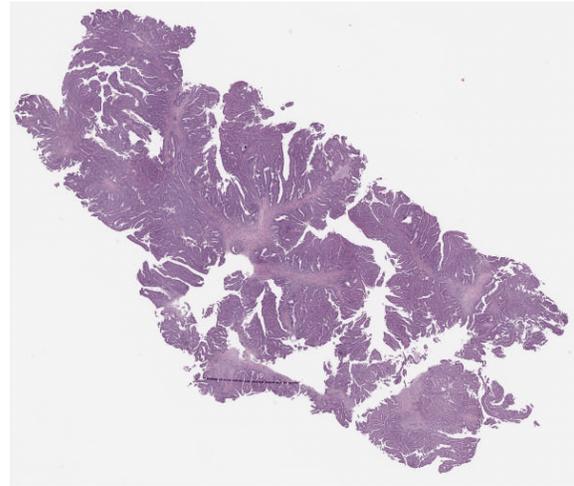


Figura 4.5: Ejemplo de una biopsia

Es por esto que las imágenes del CPTAC necesitan un umbral de saturación y filtro de mediana más grande y un tamaño del elemento estructurante del cierre morfológico más pequeño. En cuanto a las imágenes independientes, algunas son resecciones y otras biopsias. De todos modos, la mayor diferencia es que la umbralización de la saturación no es suficiente para segmentar el tejido en dichas imágenes. Esto se debe al tipo de digitalizador que se utiliza para su digitalización. Es por esto que para segmentar el tejido en las imágenes de archivo, se utiliza el método de Otsu en donde, a partir de la distribución de tonos en la imagen se elige el mejor umbral para la binarización de la misma. La siguiente imagen muestra las dos formas de visualizar la segmentación del tejido. Por un lado, la imagen de arriba muestra una máscara contorneada sobre la imagen original delimitando las regiones de tejido. Por otro lado, la imagen de abajo muestra la máscara aplicada sobre la imagen original, es decir, se elimina todo el fondo y se mantiene únicamente la región del tejido. El contorno aproximado del tejido es almacenado para el procesamiento posterior.

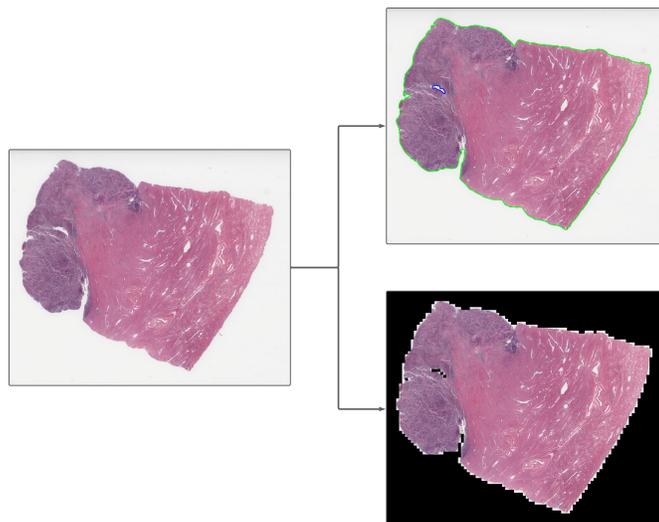


Figura 4.6: Segmentación del tejido. Obtención de la máscara y aplicación de la misma.

4.4.2. Obtención de parches

Luego de la segmentación, para cada diapositiva, el algoritmo recorta parches de 256x256 píxeles dentro de la región contorneada en el paso anterior. Esto puede observarse en la Figura 4.7. El algoritmo guarda las coordenadas de dichos parches para poder ingresarlos al siguiente paso del flujo de trabajo.

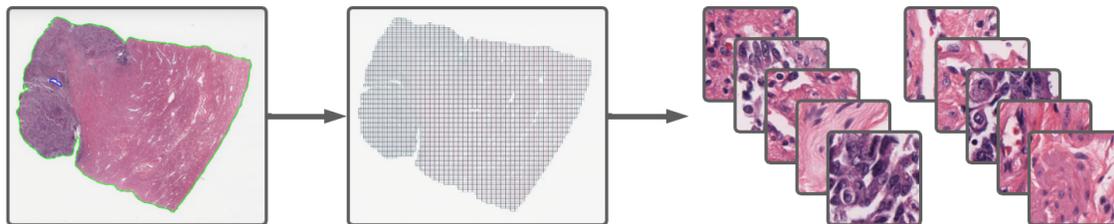


Figura 4.7: Obtención de parches a partir de la región de tejido de la diapositiva

Como se mencionó previamente, los conjuntos de datos utilizados contienen imágenes con diferentes magnificaciones. En este paso es donde se determina cuál es la magnificación a la cual se obtienen los parches y, por lo tanto, donde radica la diferencia entre el modelo entrenado en 10x y aquel entrenado en 20x. En ambos casos los parches tienen el mismo tamaño, 256x256; la diferencia está en la cantidad de tejido que contiene cada uno. Lógicamente un parche a una magnificación de 10x contiene aproximadamente cuatro veces más tejido que aquel con una magnificación

de 20x. Es por esto que la cantidad de parches en 10x es alrededor de cuatro veces menor que la cantidad en 20x. Para el caso de la extracción de parches en 20x, se obtienen en promedio 12.763 parches por imagen mientras que para aquellos extraídos en 10x, 3.249 parches por imagen. La siguiente imagen muestra la diferencia entre parches extraídos a 10x y 20x en donde se puede observar la diferencia entre la cantidad de tejido que abarca cada uno.

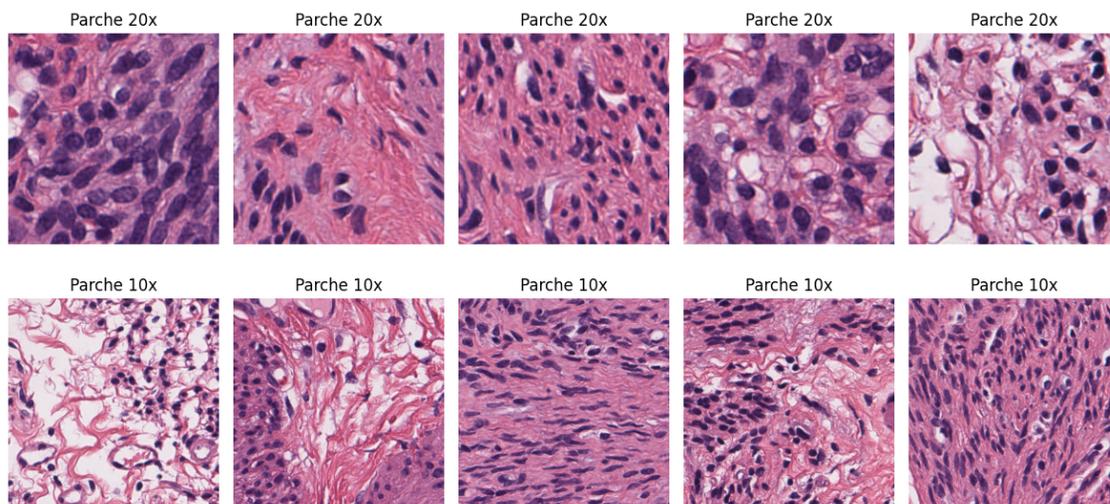


Figura 4.8: Comparación entre los parches extraídos a 10 y 20x

Es importante destacar que debido a la menor cantidad de parches extraídos a una magnificación de 10x, la velocidad de los pasos siguientes en el flujo de trabajo se reduce drásticamente, así como también se reduce el costo computacional.

4.4.3. Extracción de características

Luego de la extracción de parches de la imagen, se utiliza una red neuronal convolucional profunda para calcular una representación de características de baja dimensionalidad para cada parche de cada imagen. Esto es posible mediante la técnica de aprendizaje por transferencia en donde se utiliza un codificador de CNN con parámetros pre-entrenados. El uso de un codificador pre-entrenado no solo simplifica el proceso, sino que también acelera significativamente el entrenamiento del modelo. Esto es especialmente beneficioso en términos de tiempo y recursos computacionales.

Concretamente, se utiliza un modelo de ResNet-50 pre-entrenado en ImageNet47. Este set de imágenes es un subconjunto de ImageNet, un conjunto de datos de

imágenes que se utiliza ampliamente en el aprendizaje profundo. Es el conjunto de datos por excelencia en la actualidad para evaluar algoritmos de clasificación, localización y reconocimiento de imágenes [80]. Se utiliza este modelo agregándole una capa de *adaptive mean-spatial pooling* luego del tercer bloque residual para convertir cada parche de 256x256 píxeles en un vector de características con 1024 dimensiones. Es decir, esta capa extra comprime la información relevante del parche en tan solo un vector de 1024 elementos.

Por lo tanto, utilizando una CNN para la extracción de características, se convierten todos los parches de tejido en conjuntos de características de baja dimensionalidad en lugar de ser conjuntos de píxeles de alta dimensionalidad. De este modo, el volumen de espacio de datos se reduce casi 200 veces disminuyendo drásticamente el tiempo y el costo de los cálculos posteriores requeridos para entrenar a los modelos de clasificación. Es por esto que los modelos de aprendizaje profundo así como las inferencias de los mismos se realizan a partir de estos vectores de características y no a partir de los parches “crudos”.

4.4.4. Proceso de aprendizaje

Clasificador a nivel de imagen

En la primera tarea de aprendizaje, la red de atención genera dos conjuntos distintos de puntajes de atención, cada uno correspondiente a una de las clases posibles, es decir, MSI-H y MSS. Cada conjunto asigna un puntaje de atención a cada parche en la imagen, indicando cuánta importancia se le debe dar a ese parche para clasificar la imagen en esa clase específica. Estos puntajes de atención facilitan que la red aprenda las características morfológicas cruciales para cada clase, diferenciando claramente la evidencia positiva de la negativa. La red crea conjuntos separados para cada clase, combinando y resumiendo las representaciones de cada parche para formar una representación general a nivel de diapositiva. A continuación se explica este proceso de manera más exhaustiva y se puede ver esquematizado en la Figura 4.9.

El modelo cuenta con una red de atención (*attention network*) compuesta de varias capas totalmente conectadas, por las cuales pasa una imagen a la vez. La primera sección de estas capas es un núcleo de atención compartido entre las dos clases posibles. La entrada a este núcleo son los k vectores de 1024 elementos obtenidos a partir de la CNN en el paso de extracción de características. Siendo k el número

de parches en la imagen. Este es el paso número (1) de la Figura 4.9. La primera capa del núcleo de atención comprime a cada vector de características para obtener la misma cantidad de vectores, pero de 512 elementos.

Luego, la red se divide en dos ramas de atención paralelas, una para cada clase. Esto se visualiza en el paso número (2) de la Figura 4.9. Dentro de cada rama la primera capa es la responsable de calcular un puntaje de atención para cada parche. Éste indica cuánto contribuye el parche en la decisión de clasificar a la imagen completa dentro de la clase asociada a esa rama. Esto se visualiza en el paso número (3) de la Figura 4.9. La fórmula concreta para el cálculo del puntaje de atención se encuentra en la sección 9.3 del anexo.

La siguiente capa dentro de cada rama tiene como objetivo generar una representación a nivel de la imagen que tenga en cuenta las características de cada parche considerando su puntaje de atención. La representación para cada clase se logra sumando las características de cada parche, las cuales están ponderadas por su respectivo puntaje de atención. De esta forma se obtiene un único vector de 512 elementos que “resume” toda la información relevante de la imagen para cada clase. Esto se esquematiza en el paso número (4) de la Figura 4.9.

Finalmente, la última capa de cada rama es la capa de clasificación en donde se computa un puntaje para cada clase a partir del vector de representación obtenido en la capa anterior. Este es el paso número (5) de la Figura 4.9.

Por lo tanto, al final de la red de atención, se tienen dos puntajes de predicción, uno para cada clase. A estos puntajes se les aplica una función de activación *softmax* para obtener una distribución de probabilidad que indica la probabilidad de que la imagen pertenezca a cada una de las clases. De este modo, la salida del modelo comprende dos probabilidades: una para la clasificación como MSI-H y otra para MSS. Es importante destacar que estas probabilidades se suman a 1, proporcionando una distribución completa de la predicción. Este último paso corresponde al número (6) en la Figura 4.9.

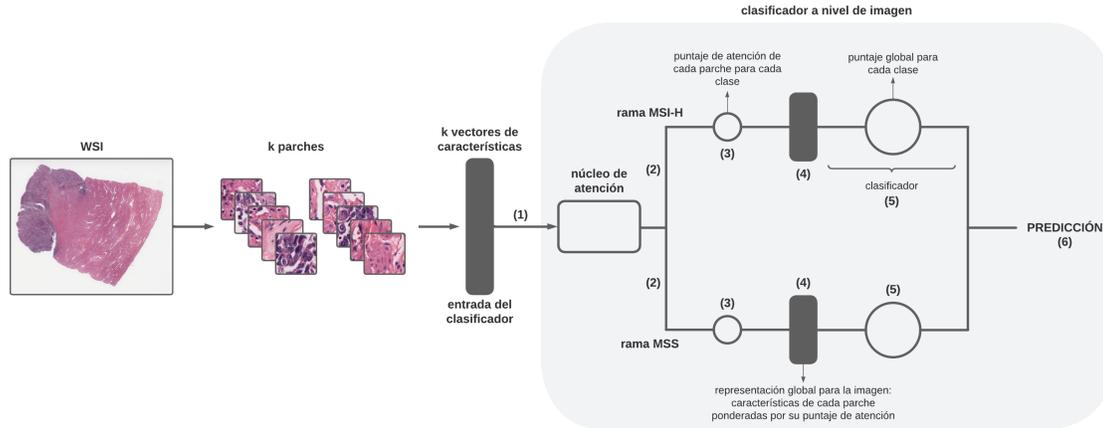


Figura 4.9: Proceso de aprendizaje del clasificador a nivel de imagen a través de la red de atención

Clasificador a nivel de instancia

Para fomentar aún más el aprendizaje de características específicas para cada clase, se introduce un *clustering* binario. El objetivo de este *clustering* es aprender de los parches, es decir, las instancias, con mayor y menor puntaje de atención para cada clase. La idea general es, para cada clase, agrupar y separar los parches de la imagen que son evidencia positiva de aquellos que son evidencia negativa. En vez de agrupar todos los parches de la imagen, se agrupan únicamente aquellos que corresponden a las regiones con mayor y menor puntaje de atención, es decir, las regiones más relevantes para la clasificación.

La forma en la que se aplica este mecanismo es introduciendo, para cada clase, una capa totalmente conectada luego de la primera capa del núcleo de atención explicado en la sección anterior. A estas capas se las denomina capas de agrupamiento. La entrada de cada una de las capas es el vector de 512 elementos obtenido a partir de la compresión del vector de características. Por otro lado, la salida de cada una de ellas es un vector de dos elementos que contiene los puntajes de asignación de *cluster* o puntaje de asignación de grupo. Para cada clase, estos puntajes reflejan si dicho parche se considera como evidencia positiva o negativa para la clase. Los pesos de esta capa se actualizan luego de cada iteración y la forma en la que se supervisa este agrupamiento es mediante la generación de pseudo-etiquetas que se obtienen a partir de la red de atención explicada en la sección anterior. A continuación se explica el modo en el que se generan dichas pseudo-etiquetas.

Dado que se conoce la clase para una determinada imagen, MSI-H o MSS, para cada imagen hay una rama de la red que corresponde a la clase correcta, “rama-correcta”, y una que corresponde a la incorrecta, “rama-incorrecta”. Por ejemplo, en el caso de una imagen etiquetada como MSI-H, la rama correspondiente a los puntajes de atención para la clase MSI-H es la “rama-correcta” mientras que la que corresponde a la clase MSS es la “rama-incorrecta”. En el caso de la “rama-correcta”, se arma una lista en orden ascendente con todos los puntajes de atención para cada parche. A los 8 parches con puntajes más altos se les asigna la etiqueta de agrupamiento positiva, ya que se los considera como evidencia fuerte para la clase correcta. Por otro lado, a los 8 parches con menor puntaje se les asigna la etiqueta de agrupamiento negativa y se los considera como evidencia fuerte negativa para la clase correcta. Se utilizan los 8 parches con mayor y menor puntaje ya que los autores de CLAM han demostrado que este número es el que muestra mejores resultados y que incluso la variación del mismo no genera cambios significativos en el desempeño del modelo.

Para las tareas de clasificación entre distintos tipos de tumores, se asume que las clases son mutuamente excluyentes. Es decir que, si el tumor es inestable, no va a ser estable y viceversa. Por lo tanto se impone una supervisión extra en la “rama-incorrecta”. A los 8 parches con puntaje más alto en la “rama-incorrecta” se les asigna también la etiqueta de agrupamiento negativa ya que se considera a los mismos como falsa evidencia positiva.

Por lo tanto, se compara la salida de la capa de agrupamiento con la pseudo-etiqueta y de este modo se entrena y mejora el desempeño del modelo. La siguiente figura muestra un esquema del funcionamiento de la capa de agrupamiento.

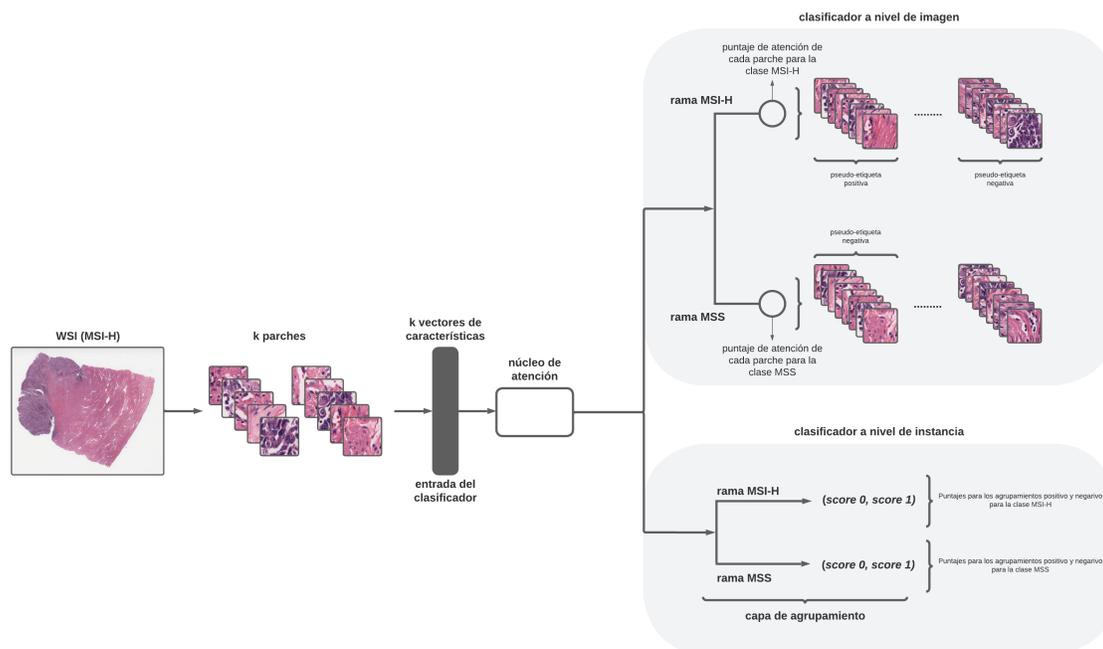


Figura 4.10: Funcionamiento de la capa de agrupamiento ejemplificada para una imagen MSI-H

Cálculo de la función de pérdida

Los pesos de las conexiones entre neuronas se inicializan aleatoriamente y se van actualizando en cada iteración con el fin de minimizar la función de pérdida. Para cada iteración se analiza el nivel de aprendizaje adquirido mediante dos funciones de pérdida, una para la clasificación a nivel de imagen y otra para la clasificación a nivel de instancia. Para el clasificador a nivel de imagen, se utiliza la función de pérdida de entropía cruzada, la cual compara la distribución de probabilidad obtenida para la imagen con la etiqueta real de la misma. En cambio, para el clasificador a nivel de instancia, se utiliza la función de pérdida SVM binaria suavizada. En esta función se comparan los puntajes de cada parche con los valores de su pseudo-etiqueta.

Las dos funciones de pérdida son combinadas para obtener una única función de pérdida representativa de la iteración de aprendizaje. Esto se logra ponderando las funciones con escalares y sumándolas como se puede observar en la ecuación (4.1).

$$L_{total} = 0,7 \cdot L_{imagen} + 0,3 \cdot L_{instancia} \quad (4.1)$$

Al igual que con el resto de los hiperparámetros, se eligieron estas ponderaciones ya que las mismas resultaron en un mejor desempeño para las tres tareas de clasificación

de CLAM. Cabe destacar que estos parámetros se establecen de tal manera que se le atribuye mayor relevancia a la clasificación a nivel de imagen que a la clasificación a nivel de instancia.

El proceso de aprendizaje cuenta con múltiples iteraciones en las cuales se busca disminuir la función de pérdida mencionada. Esto se hace a través de un optimizador de Adam. La fórmula y los parámetros de este optimizador se encuentran detallados en la sección 9.4 del anexo.

Una vez que los pesos fueron actualizados, se lleva a cabo una siguiente iteración en la que se repite todo el proceso descrito. El modelo realiza como mínimo 50 iteraciones, y como máximo 200. En este modelo se utiliza validación cruzada, la cual monitorea la función de pérdida en el conjunto de validación. Cuenta con un criterio de terminación temprana, que indica que si la función de pérdida para el conjunto de validación no disminuye por 20 iteraciones consecutivas, el entrenamiento debe frenar. Una vez terminado el entrenamiento, ya sea por la terminación temprana o por alcanzar el máximo de iteraciones, se guarda la iteración con la menor función de pérdida como el modelo final. Este modelo es luego evaluado en el conjunto de evaluación.

En esta arquitectura se utiliza validación cruzada de Monte Carlo de 10 instancias. Esto significa que la partición del conjunto de datos en 80 % entrenamiento, 10 % validación y 10 % evaluación se lleva a cabo 10 veces, en las que la selección de qué imágenes van en qué clase es aleatoria. De esta manera, el proceso de entrenamiento, validación y evaluación se lleva a cabo 10 veces, generando 10 modelos distintos. En el caso de este trabajo, se selecciona finalmente el modelo con el mejor desempeño.

4.5. *Heatmaps* como herramienta visual

CLAM, mediante el aprendizaje basado en atención, prescinde de etiquetas para cada parche y, en cambio, genera puntuaciones de atención para evaluar la relevancia de cada uno en la clasificación de la imagen para una clase específica. Estas puntuaciones pueden transformarse en *heatmaps*, es decir, representaciones visuales del tejido coloreadas según la importancia asignada. Al normalizar las puntuaciones y convertirlas en colores RGB, se destacan los parches de alta atención en rojo, indicando su fuerte contribución a la predicción. Por otro lado, los parches de baja atención se muestran en azul, señalando una menor influencia en la predicción del

modelo. Este enfoque facilita la identificación visual e interpretación de las áreas clave en la imagen. Se utiliza esta herramienta no sólo para identificar los parches más relevantes para cada clase, sino también para una tarea adicional: localizar el tumor dentro de las imágenes.

4.5.1. Análisis de los parches más relevantes para cada clase

El objetivo fundamental del modelo radica en su capacidad para aprender y reconocer las características morfológicas distintivas que definen cada clase de interés. Con este fin, luego de seleccionar al modelo con el mejor desempeño, se analizan los *heatmaps* de las imágenes clasificadas por el algoritmo para entender cuáles son las áreas donde se centra la atención al categorizar una imagen en una clase específica.

Por lo tanto, como parte de la metodología, se eligen aleatoriamente diez imágenes correctamente clasificadas para cada clase. Para cada una de estas imágenes, se identifican los cinco parches con el mayor puntaje de atención asignado a la clase correcta. Estos parches se someten posteriormente al análisis de una patóloga experta, con el propósito de evaluar si existen características biológicas perceptibles por el ojo humano que sean distintivas para cada clase.

4.5.2. Localización del tumor

Para agregar valor a la herramienta de patología digital, se busca también la localización del tejido tumoral dentro de la WSI. En lugar de entrenar al modelo para llevar a cabo esta tarea, se aprovechó uno de los modelos que entrenaron y validaron los autores de CLAM.

Como se mencionó anteriormente, los creadores de CLAM emplearon los conjuntos de datos CAMELYON16 y CAMELYON17 de manera conjunta, generando un conjunto de datos de 899 imágenes para abordar la tarea de detección de metástasis de cáncer de mama en ganglios linfáticos. Este conjunto incluye 591 imágenes negativas (sin metástasis) y 308 positivas (con metástasis), todas capturadas a una magnificación de 40x y con una resolución de 0,25 mpp. En promedio, se generan 41.802 parches por imagen. Los resultados obtenidos utilizando el proceso de aprendizaje detallado anteriormente fueron destacados, con un AUROC promedio de $0,953 \pm 0,029$ en los 10 pliegues de entrenamiento/validación/evaluación.

Adicionalmente, se llevó a cabo una evaluación de este modelo utilizando un conjunto de datos independiente. Estas muestras fueron recopiladas internamente

en el BWH, abarcando archivos histológicos desde 2016 hasta 2019. Tras excluir las imágenes con tejido dañado o notables marcas sobre el mismo, se conformó un conjunto de datos de 133 imágenes (66 negativas y 67 positivas). Estas imágenes, con una magnificación de 40x y una resolución de 0,25 mpp, presentan un promedio de 51.426 parches por imagen. Para este conjunto de datos independiente, los resultados son muy alentadores, exhibiendo un AUROC promedio de $0,940 \pm 0,015$.

Basándose en estos resultados, uno de los modelos generados durante este proceso de entrenamiento se descargó desde la plataforma de CLAM¹. Siguiendo la misma metodología de segmentación de tejido, obtención de parches y extracción de características, se procedió a utilizar dicho modelo para llevar a cabo la localización del tumor en las muestras de carcinomas endometriales incluidas en este trabajo. Es relevante señalar que, debido a que el modelo fue entrenado con imágenes en 40x, éste es aplicable únicamente para la detección de tumores en imágenes con la misma magnificación. Esta limitación representa una consideración importante en cuanto a la utilidad práctica de la herramienta, dado que un porcentaje de las imágenes del conjunto de datos empleado en este estudio presenta una magnificación de 20x.

Aunque se consideró entrenar al modelo con muestras de carcinoma endometrial para su aplicación en imágenes en 20x, la insuficiente disponibilidad de muestras de tejido sano impidió que se llevara a cabo dicho entrenamiento. Además, este aspecto no constituye el enfoque central de nuestra investigación.

Debido a que el modelo utilizado en esta tarea fue entrenado con tejidos de otro órgano, se procuró la validación de una patóloga para asegurar el correcto desempeño. En la Figura 4.11 se presenta una comparación entre el tejido tumoral marcado por la patóloga y los *heatmaps* generados por CLAM, donde los parches con un tono de rojo más intenso indican la presencia de tejido tumoral.

¹<https://drive.google.com/drive/folders/1GDpYFs9QUoBBJhupD0E11MlfsEq9YMi>

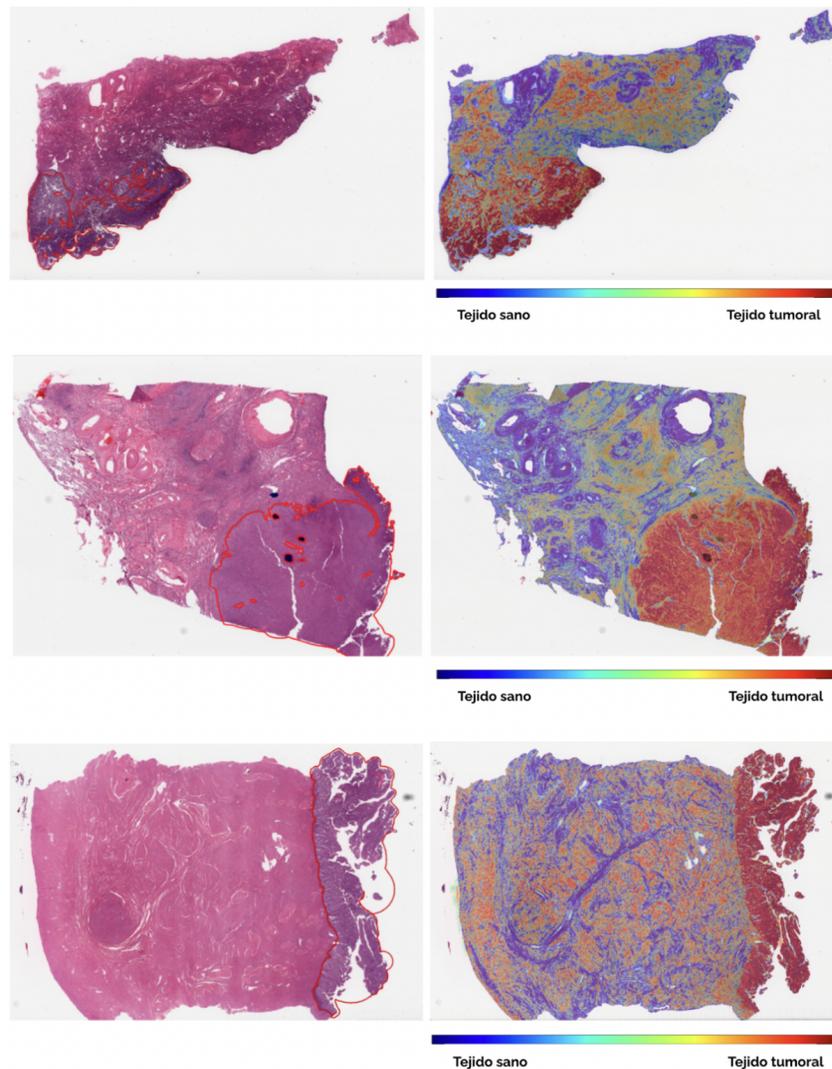


Figura 4.11: Comparación entre marcaciones de tumor realizadas por una patóloga y *heatmaps* para la localización del tejido tumoral

Los resultados fueron muy satisfactorios, salvo en los casos en donde las muestras presentan un alto nivel de infiltración linfocitaria. El mismo es tejido sano que en ciertos casos el algoritmo CLAM reconoce como tejido tumoral. En la Figura 4.12 se puede observar la comparación entre el tejido marcado y el *heatmap* en el caso de un caso que presenta un alto nivel de infiltración linfocitaria. La infiltración linfocitaria se encuentra marcada en amarillo en la imagen de la izquierda. En la imagen de la derecha se observa cómo el algoritmo percibe dicha zona como tumoral.

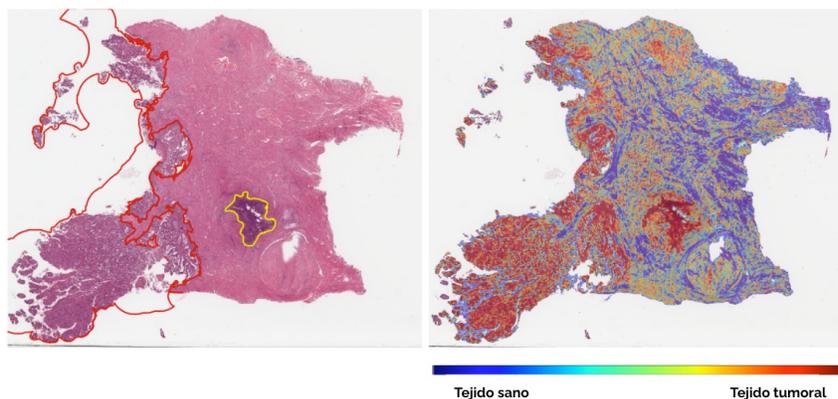


Figura 4.12: Comparación entre marcaciones de tumor realizadas por una patóloga y *heatmaps* para la localización del tejido tumoral para el caso de una infiltración linfocitaria (marcación amarilla)

4.6. Exploración de otras líneas de trabajo

Durante el transcurso de la investigación se desarrollaron dos líneas de trabajo en paralelo. Por un lado se exploró la alternativa de utilizar la arquitectura CLAM tal como se explicó en todas las secciones previas. Por otro lado, también se exploró la alternativa de utilizar *Vision Transformers* para desarrollar el modelo de clasificación. Con el objetivo de limitar la extensión del trabajo a los tiempos acordes a un proyecto final de carrera se continuó con la alternativa de CLAM, ya que la misma fue la primera en producir muy buenos resultados. Sin embargo, es necesario ahondar en la segunda alternativa, ya que resulta muy prometedora para futuras direcciones de esta investigación.

Esta alternativa consiste en utilizar los parches tumorales de las diapositivas para entrenar a un modelo pre-entrenado de ViT. La principal diferencia con la alternativa CLAM es que, para el caso de los ViTs, se necesita tener previamente detectado cuáles son los parches tumorales de la imagen. Una vez que se tienen dichos parches, se entrena a un modelo de ViT para poder detectar la inestabilidad microsatelital al nivel de un parche. Por lo tanto, cuando se quiere analizar el estado microsatelital de la imagen completa, se debe predecir el estado de cada uno de los parches para luego realizar un promedio de todos ellos y determinar el estado a nivel de la diapositiva.

Para evitar que un patólogo tenga que anotar manualmente más de 600 imágenes, se utilizó la herramienta del *heatmap* tumoral mencionada en la sección anterior. En

este caso, no sólo se generaron los *heatmaps* tumorales de las imágenes sino que se guardaron aquellos parches con un puntaje de atención tumoral superior a 0,99. Estos parches son los que luego se ingresaron al modelo de ViT para entrenarlo.

A diferencia de la arquitectura CLAM en donde se entrena el modelo con características extraídas de los parches, el modelo de ViT se entrena con los propios parches volviéndose drásticamente más costoso a nivel computacional. Es importante destacar que antes de ingresar los parches al modelo, los mismos se sometieron a una normalización utilizando el método Macenko [81] para eliminar las variaciones que surgen debido a las diferentes tinciones de las imágenes. Para el caso de CLAM, la normalización se lleva a cabo en el proceso de extracción de características y por lo tanto no es algo que se debe realizar manualmente. La siguiente imagen muestra el flujo de trabajo de la alternativa ViT.

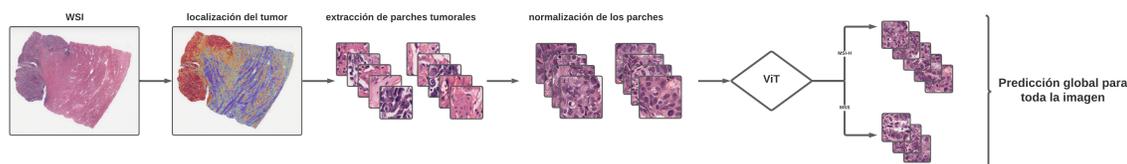


Figura 4.13: Flujo de trabajo para la utilización de ViTs como modelo de clasificación

Debido al mayor costo computacional de este método, los tiempos de procesamiento fueron mucho mayores a aquellos de CLAM, es por esto que aún no se obtuvieron resultados concretos para este método. En primer lugar, se obtuvieron todos los parches tumorales de las imágenes con una magnificación nativa de 40x del conjunto de datos de TCGA-CPTAC. Se utilizaron estos parches para luego entrenar y evaluar al modelo “vit-base-patch16-224-in21k” [82] pre-entrenado con ImageNet. Sin embargo, al llegar a la fase de realizar el ajuste fino del modelo, ya se habían alcanzado resultados destacados con la arquitectura CLAM, por lo que se optó por pausar esta alternativa. Se llevó a cabo el entrenamiento y la evaluación utilizando solo un pliegue, es decir, una única división (entrenamiento/validación/evaluación) del conjunto de datos y con un total de diez iteraciones de entrenamiento.

Si bien los resultados que se obtuvieron fueron inconclusos se describe brevemente esta alternativa por lo prometedora que es para futuras mejoras de esta herramienta. Un beneficio significativo de esta técnica por sobre CLAM es que, al recibir sólo los parches de la región tumoral, se reduce el ruido y el algoritmo se enfoca únicamente en las regiones de interés. Hoy en día se están utilizando estos modelos para la

clasificación de imágenes histológicas gracias a su capacidad para capturar patrones complejos y contextuales en este tipo de imágenes mejorando la precisión de la clasificación y permitiendo una interpretación más profunda de las características histológicas [83].

4.7. Desarrollo de un prototipo de interfaz gráfica

El propósito de este trabajo consiste en desarrollar una herramienta de patología digital que permita la evaluación de la presencia de inestabilidad microsatelital en imágenes digitalizadas de tejido tumoral dentro de una institución médica. Para alcanzar este objetivo es necesario esbozar algún tipo de plataforma que contenga al modelo entrenado y que sea interpretable y operable por un profesional de salud. En lo que compete a este trabajo, se desarrolló una interfaz gráfica que opera como un prototipo de lo que sería una aplicación o página web que se pueda acceder libre y gratuitamente desde cualquier institución médica del país. Se utilizó la aplicación QTDesigner para diseñar una interfaz gráfica que luego se exportó a python utilizando la librería PyQt5. La misma puede observarse en la sección de resultados.

Resultados

A continuación se muestran los resultados de los cuatro modelos planteados para la tarea de clasificación entre imágenes MSI-H y MSS. Estos resultados se obtuvieron al evaluar dichos modelos con el conjunto de evaluación del TCGA-CPTAC. Una vez seleccionado el modelo con mejor desempeño, el mismo se evaluó sobre el conjunto independiente.

Tal como se muestra en la Tabla 4.2, el conjunto de evaluación para los modelos desbalanceados se compone de 50 imágenes, 16 MSI-H y 34 MSS. Por otro lado, los modelos balanceados se evalúan en un conjunto de 38 imágenes, 16 MSI-H y 22 MSS. La Tabla 5.1 muestra las métricas más importantes para cada uno de los modelos y sus matrices de confusión pueden observarse en la Figura 9.2 del anexo.

	Exactitud	Precisión	Sensibilidad	AUROC	Especificidad	NPV
modelo_desbalanceado_20x	0,76	0,70	0,44	0,79	0,91	0,76
modelo_balanceado_20x	0,66	1,00	0,19	0,82	1,00	0,63
modelo_desbalanceado_10x	0,76	0,61	0,69	0,80	0,79	0,84
modelo_balanceado_10x	0,76	0,71	0,75	0,92	0,77	0,81

Tabla 5.1: Métricas para evaluar el desempeño de los distintos modelos

A la hora de analizar estos resultados, es importante considerar qué rol puede tomar este algoritmo en un laboratorio de patología. El modelo tiene dos formas distintas de equivocarse, predecir un tumor estable como inestable o un tumor inestable como estable. Las implicaciones patológicas de estos dos errores no son las mismas.

En la actualidad, en un laboratorio de patología se lleva a cabo el diagnóstico de la integridad o daño del sistema de reparación, subrogante de la inestabilidad microsatelital, mediante la técnica de inmunohistoquímica. En el caso de que la muestra resulte inestable (MMRd), se realiza una segunda prueba de biología molecular para confirmar el diagnóstico. Por el contrario, en el caso de que la muestra sea estable

(MMRp), no se realiza una segunda prueba y se siguen las vías tradicionales de tratamiento para tumores estables.

La implementación de esta herramienta digital en el laboratorio sigue la misma lógica. Si se considera la salida del algoritmo MSI-H como positiva y la MSS como negativa, en el caso de que el algoritmo clasifique un tumor estable como inestable -es decir, un falso positivo- el laboratorio va a realizar pruebas para confirmar el diagnóstico y se va a poder detectar el error.

En cambio, en el caso de que un tumor inestable sea predicho como estable, no se le realizan pruebas subsiguientes a la muestra y se “pierde” el caso. Es decir que, en el caso de un falso negativo, no hay una oportunidad de corregir el error.

Por lo tanto, a la hora de determinar qué modelo da los mejores resultados, se busca reducir la cantidad de falsos negativos, es decir, reducir la cantidad de casos MSI-H que el algoritmo predice como MSS. Esto se puede hacer al considerar principalmente la métrica de sensibilidad. La misma representa la tasa de verdaderos positivos sobre el total de los casos realmente positivos. Cuanto menor sea el número de falsos negativos, mayor será la sensibilidad.

Cuando se busca priorizar cierta métrica por sobre otras, se utiliza la técnica de la umbralización para cambiar el umbral de decisión del modelo y favorecer la clasificación de una clase por sobre la otra. Para comprender en qué consiste dicha técnica hay que tener muy en claro cuál es el formato de la salida del clasificador. Si bien puede parecer que el clasificador provee una salida binaria, es decir, MSI-H o MSS, la salida es un conjunto de dos probabilidades; la probabilidad de que la muestra sea MSI-H y la probabilidad de que sea MSS. Lógicamente, la suma de ambas probabilidades es 1. Por lo tanto, es muy común que la decisión del modelo se incline hacia la clase que presente una mayor probabilidad, es decir, que para determinar que una muestra sea MSI-H, la probabilidad de pertenecer a dicha clase deba ser mayor a 0,5. Los resultados que se muestran en la Tabla 5.1 corresponden a dicho umbral de decisión.

Para elegir el umbral de decisión que favorezca la sensibilidad sin comprometer demasiado la precisión, se utilizó la curva PR. Como se mencionó previamente, dicha curva es una representación gráfica de la precisión frente a la sensibilidad para cada umbral de decisión.

Para elegir cuál de los cuatro modelos es el que presenta un mejor desempeño con el conjunto de datos TCGA-CPTAC, se buscó el umbral óptimo para cada caso

y se compararon sus métricas. A continuación se muestra el procedimiento para el modelo seleccionado, es decir, aquel que presenta un mejor desempeño para esta tarea en particular. Dicho modelo resulta ser el modelo_balanceado_10x.

5.1. Selección del umbral óptimo para la tarea de clasificación

Una vez que se evalúa al modelo y se obtienen para cada muestra las probabilidades de pertenecer a cada una de las clases, se realiza la curva PR del mismo. Las Figuras 5.1 y 5.2 muestran la curva PR del modelo_balanceado_10x y una tabla con su sensibilidad y precisión para cada umbral de decisión. Es importante destacar que el umbral de decisión es el valor mínimo que debe obtener la probabilidad de pertenecer a la clase MSI-H para que la muestra sea clasificada como MSI-H.

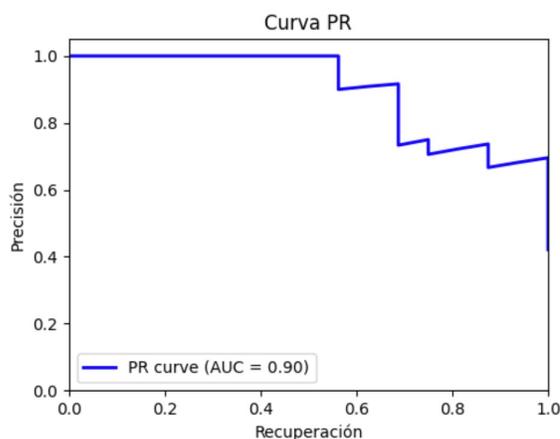


Figura 5.1: Curva PR del modelo_balanceado_10x

Tabla de métricas de la curva PR:

	Umbral PR	Precision	Recall	AUC-PR
0	0.019888	0.421053	1.0000	0.897853
1	0.073408	0.457143	1.0000	0.897853
2	0.165206	0.500000	1.0000	0.897853
3	0.194382	0.551724	1.0000	0.897853
4	0.293069	0.615385	1.0000	0.897853
5	0.324987	0.695652	1.0000	0.897853
6	0.404674	0.700000	0.8750	0.897853
7	0.502563	0.705882	0.7500	0.897853
8	0.565757	0.785714	0.6875	0.897853
9	0.627706	0.909091	0.6250	0.897853
10	0.649100	1.000000	0.5000	0.897853
11	0.763000	1.000000	0.3125	0.897853
12	0.876280	1.000000	0.1250	0.897853

Figura 5.2: Posibles umbrales de decisión para el modelo_balanceado_10x

Para este caso se toma un umbral de 0,33 ya que maximiza la sensibilidad, mientras que la precisión alcanza un valor de 0,7. En general, para el caso de varias tareas de clasificación una precisión por debajo de 0,8 no es aceptable. Sin embargo, para el caso de esta tarea en particular, en donde la sensibilidad es realmente muy importante y, por la naturaleza de la misma, el flujo de trabajo permite la posibilidad de detectar falsos positivos en un futuro cercano, se acepta una precisión de 0,7. Este procedimiento se realiza para los 4 modelos. La Tabla 5.2 muestra el valor del umbral elegido para cada uno de ellos, así como sus métricas. Por otro lado, la Figura 5.3 muestra las matrices de confusión para cada uno de ellos. Las tablas

con los distintos umbrales, así como las curvas PR para cada uno de los modelos, se pueden observar en las Figuras 9.3, 9.4 y 9.5 del anexo.

	Umbral	Exactitud	Precisión	Sensibilidad	AUROC	Especificidad	NPV
modelo_desbalanceado_20x	0.26	0,70	0,52	0,94	0,79	0,59	0,95
modelo_balanceado_20x	0.34	0,71	0,61	0,88	0,82	0,59	0,87
modelo_desbalanceado_10x	0.49	0,76	0,60	0,75	0,80	0,76	0,87
modelo_balanceado_10x	0.33	0,82	0,70	1,00	0,92	0,68	1,00

Tabla 5.2: Métricas de cada uno de los modelos luego de la umbralización

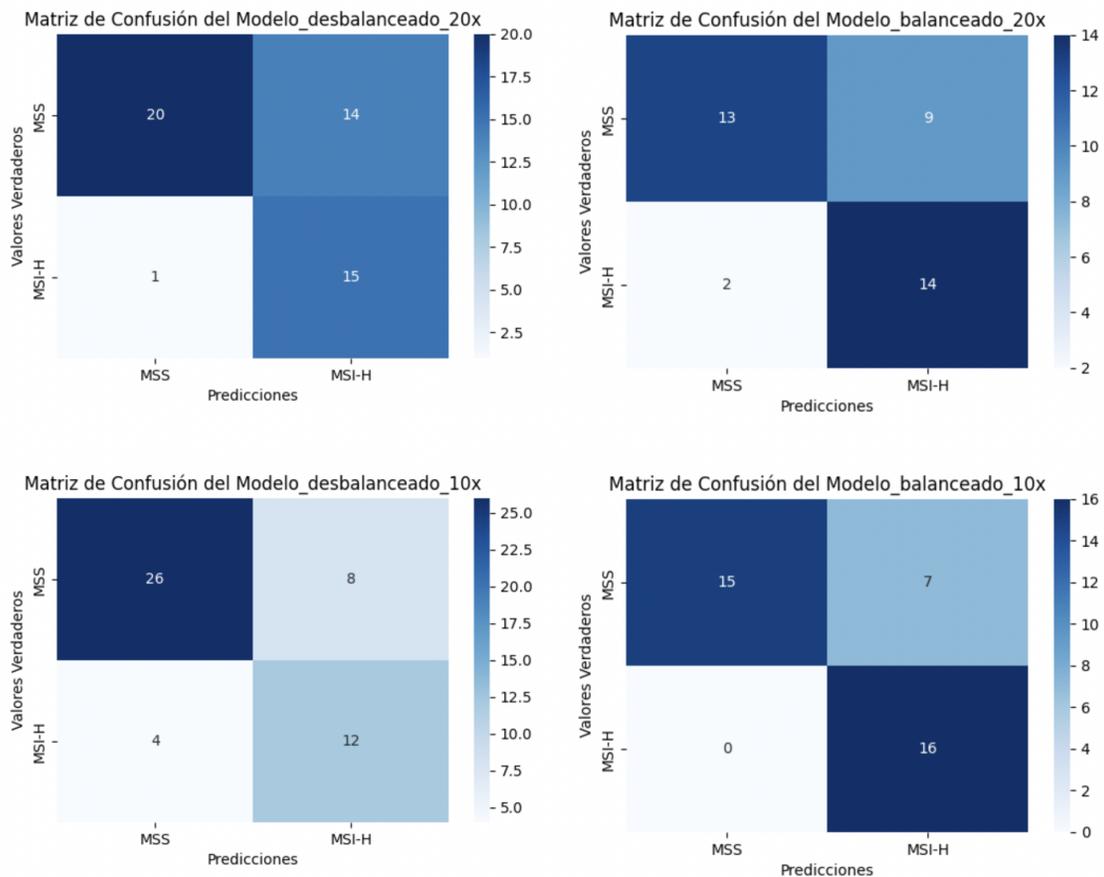


Figura 5.3: Matrices de confusión de los 4 modelos luego de la umbralización

Teniendo en cuenta los resultados obtenidos, se eligió al modelo_balanceado_10x como el que mejor se adapta a los requisitos de la tarea. A continuación se lo evaluó con el conjunto independiente.

5.2. Evaluación del modelo con el conjunto independiente

En esta sección se realiza la evaluación del modelo seleccionado con el conjunto de datos independiente. Como se mencionó anteriormente, dicho conjunto de datos contiene 51 imágenes de las cuales 22 son MSI-H y 29, MSS. La siguiente figura muestra la matriz de confusión para dicha evaluación seguida por una tabla con sus métricas más relevantes. Es importante destacar que en esta primera evaluación se utiliza el mismo umbral seleccionado como óptimo cuando se evaluó el modelo en el dataset TCGA-CPTAC.

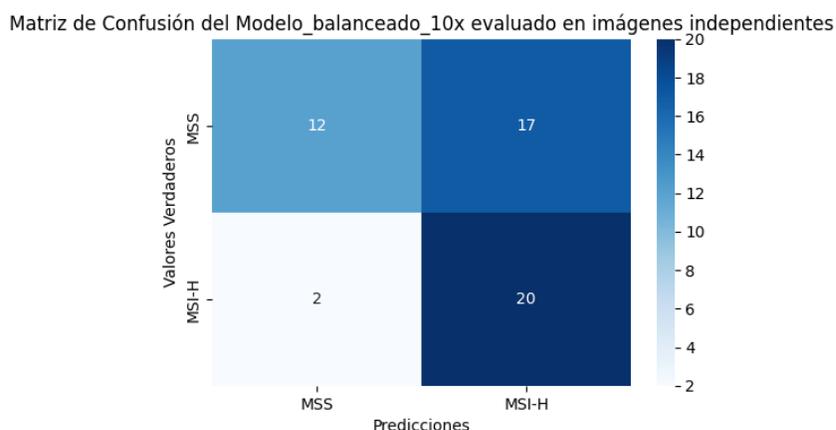


Figura 5.4: Matriz de confusión del modelo_balanceado_10x evaluado con imágenes del conjunto independiente

	Umbral	Exactitud	Precisión	Sensibilidad	AUROC	Especificidad	NPV
modelo_balanceado_10x	0,32	0,63	0,54	0,91	0,74	0,41	0,86

Tabla 5.3: Métricas del modelo_balanceado_10x evaluado en el conjunto independiente

El desempeño del modelo disminuye drásticamente con el conjunto de datos independiente con respecto al del TCGA-CPTAC. Esta pérdida significativa de precisión puede adjudicarse en gran parte a las distintas características que presentan las imágenes del conjunto independiente. Estas diferentes características se abordan en la siguiente sección pero incluyen principalmente diferencias en la coloración, fijación y digitalización de las muestras. De todos modos, una posible solución que se exploró fue modificar el umbral de decisión de tal modo que el modelo pueda predecir mejor el estado microsatelital en imágenes con características tan diferentes

a las del TCGA-CPTAC. A continuación se puede observar la curva PR del modelo_balanceado_10x evaluado en el conjunto de datos independiente junto con una tabla que muestra la sensibilidad y precisión para cada umbral de decisión.

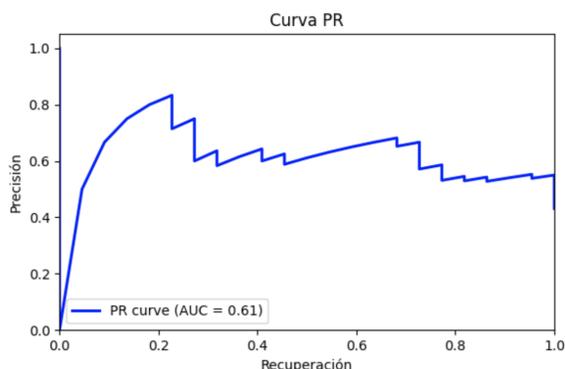


Figura 5.5: Curva PR del modelo_balanceado_10x para el conjunto independiente

Tabla de métricas de la curva PR:

	Umbral PR	Precision	Recall	AUC-PR
0	0.132395	0.431373	1.000000	0.611669
1	0.214417	0.468085	1.000000	0.611669
2	0.254194	0.511628	1.000000	0.611669
3	0.292812	0.538462	0.954545	0.611669
4	0.410620	0.542857	0.863636	0.611669
5	0.462392	0.548387	0.772727	0.611669
6	0.496512	0.592593	0.727273	0.611669
7	0.574884	0.652174	0.681818	0.611669
8	0.621909	0.631579	0.545455	0.611669
9	0.704069	0.600000	0.409091	0.611669
10	0.743458	0.636364	0.318182	0.611669
11	0.810576	0.714286	0.227273	0.611669
12	0.834257	0.666667	0.090909	0.611669

Figura 5.6: Posibles umbrales de decisión para el modelo_balanceado_10x con el conjunto independiente

Se puede observar que ningún umbral logra obtener un desempeño acorde a lo que se busca para esta tarea; si bien se puede disminuir el umbral para aumentar levemente la sensibilidad y la exactitud, el problema en este caso yace en la baja precisión y especificidad que presenta el modelo. Este problema no puede resolverse variando el umbral. En la siguiente sección se discuten las posibles mejoras que pueden hacerse tanto en el desarrollo del algoritmo como en la obtención de las muestras, para lograr un mejor desempeño del modelo en las imágenes de archivo de un laboratorio de patología.

5.3. Evaluación de la atención

Si bien hemos omitido la selección manual de los parches tumorales, la premisa subyacente es que el algoritmo pueda, a partir de una imagen SVS que abarque toda la diapositiva, dirigir su atención hacia las regiones críticas, es decir, el tumor, prescindiendo de las anotaciones de ROI o a nivel de cada parche. Hemos empleado la herramienta de los *heatmaps* para visualizar cuáles son los parches a los que el modelo seleccionado otorga un puntaje de atención más elevado para cada clase. Este proceso se ilustra en las siguientes imágenes: a la izquierda, aquellas correspondientes a la

clase inestable, y a la derecha, las relacionadas con la clase estable. Estas imágenes se contrastan con sus *heatmaps*, que delinean las regiones tumorales, con el propósito de verificar que los parches a los que se les asigna un puntaje de atención más alto para cada clase efectivamente pertenezcan a la región tumoral.

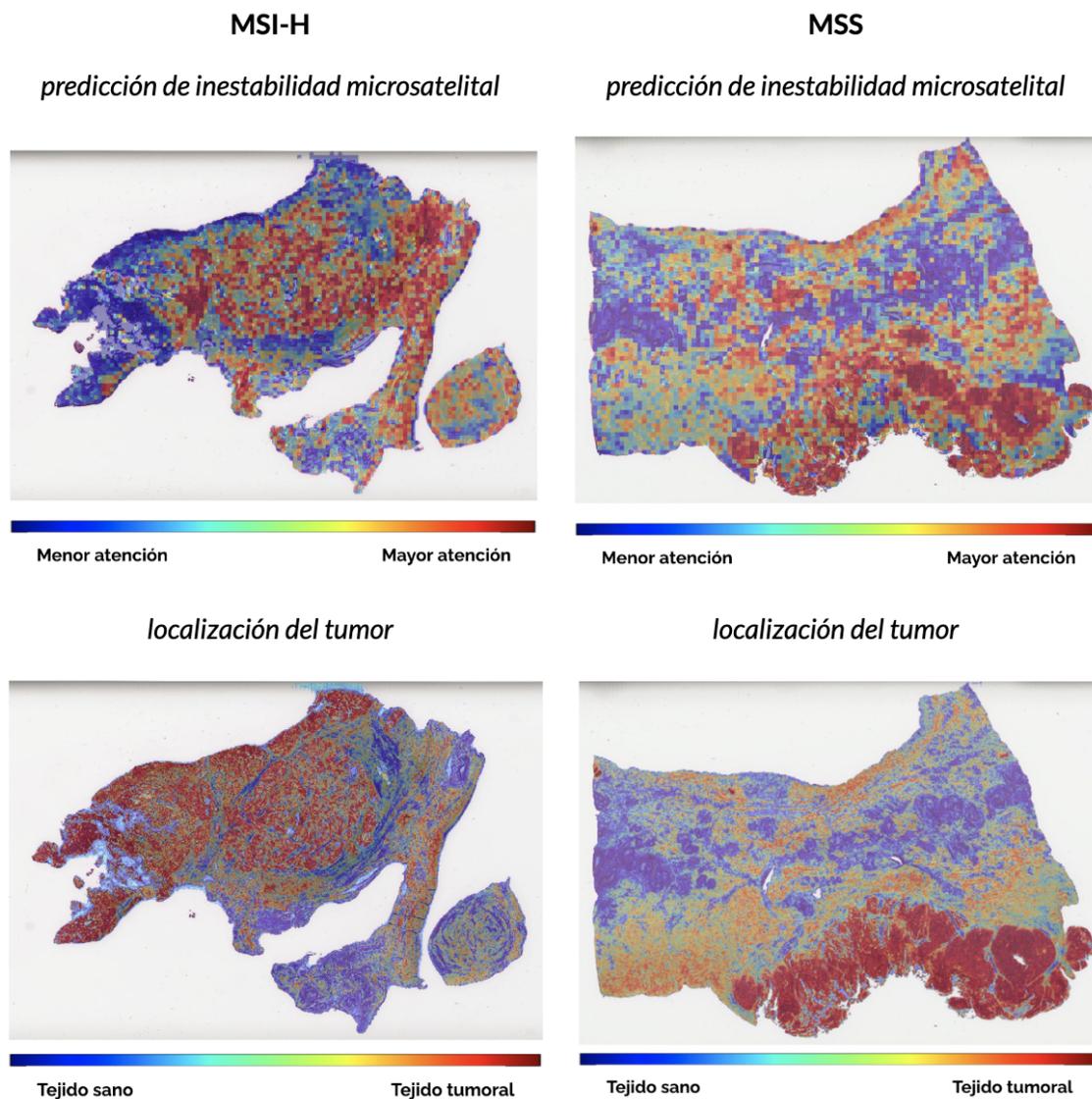


Figura 5.7: Comparación entre las regiones en las que el algoritmo pone atención y las regiones tumorales

La Figura 5.7, evidencia que la atención del modelo no se limita exclusivamente a las regiones tumorales. Esta discrepancia representa una limitación importante del

modelo, la cual se refleja coherentemente en los valores de las métricas asociadas al mismo. A continuación, se analiza con mayor detalle hacia qué áreas dirige su atención el modelo.

5.4. Interpretación biológica de los parches distintivos para cada clase

En esta sección, se presentan visualizaciones de los parches a los cuales el modelo les atribuyó un mayor puntaje de atención para cada clase. Tal como se observa en la sección anterior, tanto para la clase MSI-H como para la MSS, el algoritmo no enfoca su atención exclusivamente en la región tumoral. En la siguiente figura, se presentan algunos ejemplos de los parches con los puntajes de atención más altos para cada clase. Una patóloga experta describió a qué estructura biológica pertenecen estos parches.

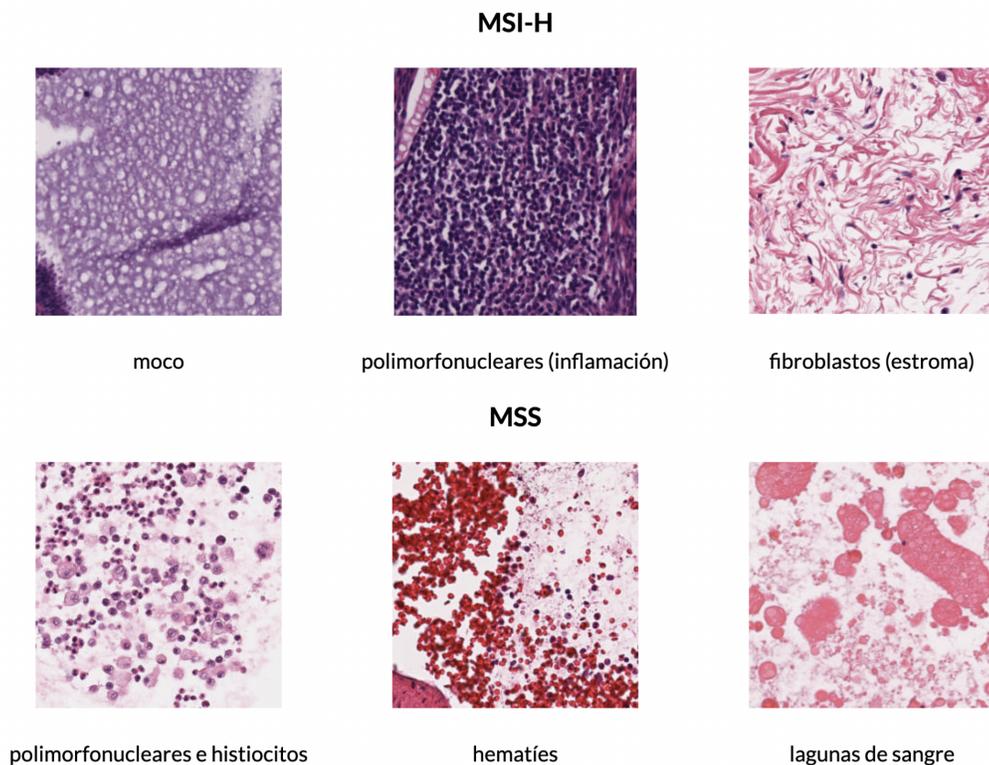


Figura 5.8: En el modelo digitalizado, parches con elevado puntaje de atención para cada clase

5.5. Interfaz gráfica

Al desarrollar la interfaz gráfica para poder implementar el modelo en una institución médica, el objetivo es su fácil implementación e interpretación por un profesional de salud. Se buscó diseñar una interfaz interactiva que contenga todos los pasos explicados en las secciones anteriores, de tal modo que cuando el usuario seleccione una imagen, la interfaz le devuelva la predicción del estado microsatelital de la misma.

En primer lugar, el usuario tiene la posibilidad de seleccionar la imagen SVS que busca clasificar, a partir de los archivos de su dispositivo. Además, el usuario debe seleccionar una de las siguientes opciones que mejor describa a la imagen: TC-GA, CPTAC, Independiente-biopsia o Independiente-resección. Esto es crucial para segmentar adecuadamente el tejido dentro de la imagen, ya que, como se mencionó en secciones anteriores, los parámetros de segmentación dependen exclusivamente de las características de la imagen. Una vez realizada esta selección, se efectúa el procesamiento de la imagen. El usuario puede, entonces, visualizar la imagen a clasificar, junto con su máscara de tejido y la segmentación del mismo.

Por otro lado, el usuario también tiene la opción de realizar la localización del tumor y visualizar el *heatmap* resultante. Este proceso es complementario, dado que no tiene relevancia en la predicción del estatus microsatelital, pero es una herramienta útil de la cual los profesionales médicos pueden sacar provecho. Esta función solo está disponible para aquellas imágenes cuya magnificación nativa es de 40x, por lo cual se incluyen controles internos que no permiten la segmentación en el caso de que la imagen tenga otra magnificación.

Por último, los resultados generados son guardados en el dispositivo del usuario y se crea una lista con el nombre de las imágenes procesadas, junto con su predicción. Esta lista queda disponible en la interfaz gráfica, de manera que el usuario pueda visualizar viejos resultados sin necesidad de volver a procesar la imagen. En la Figura 5.9 se puede observar la interfaz gráfica luego de haber procesado una imagen del tipo Independiente - resección y que fue predicha correctamente como MSI-H.

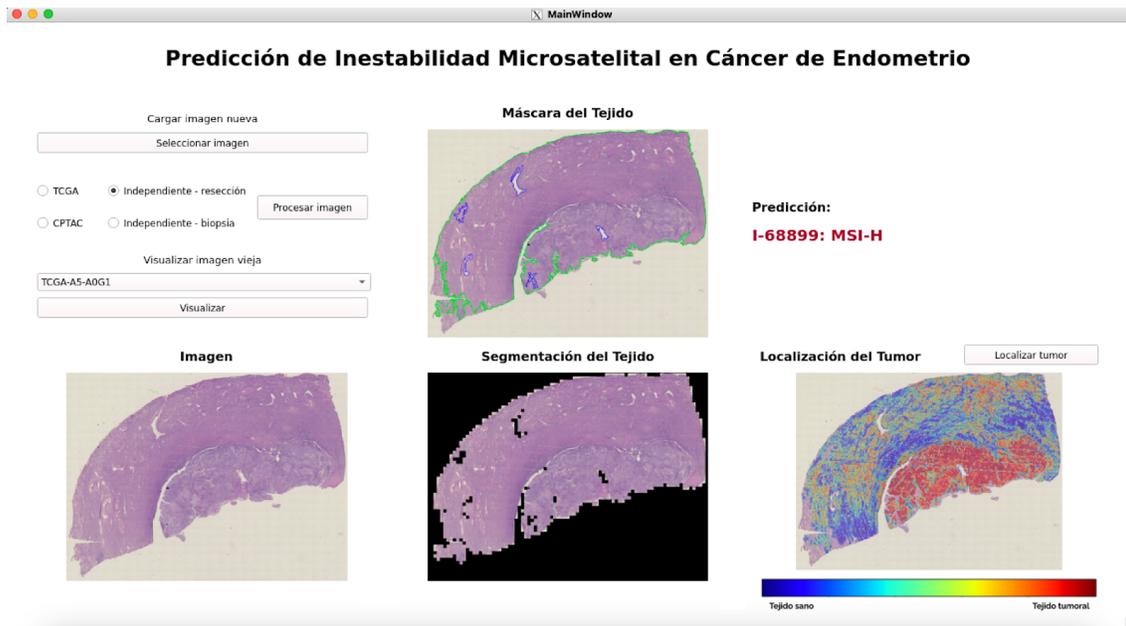


Figura 5.9: Prototipo de Interfaz Gráfica

La duración del procesamiento de la imagen, desde que se la selecciona hasta poder visualizar los resultados, se encuentra por debajo de los dos minutos. La segmentación del tejido tumoral tiene una duración de entre 5 y 8 minutos. Dado que esta segunda tarea no es necesaria para obtener la predicción del estado microsatelital, no se considera a esta segunda duración como parte del tiempo de procesamiento.

Discusión

La incorporación de modelos de inteligencia artificial en la predicción de inestabilidad microsatelital en cáncer de endometrio representa un avance significativo en el campo de la patología molecular. Los resultados obtenidos en este estudio demuestran la prometedora capacidad de nuestro modelo para desempeñar un papel fundamental como herramienta de tamizaje, tanto como complemento o como sustituto potencial de los métodos convencionales, en particular, la inmunohistoquímica.

La comparación directa entre nuestro modelo y la IHQ o la biología molecular, revela ventajas sustanciales en términos de eficiencia temporal. Si bien es cierto que los resultados actuales pueden no alcanzar una exactitud óptima, la prometedora sensibilidad del modelo sugiere su idoneidad para suplir o respaldar estas metodologías como método de primera línea en la predicción de MMRd.

Uno de los aspectos más destacados de nuestro modelo es su tiempo de predicción excepcionalmente rápido, con un límite que no supera los 2 minutos. Este contraste es particularmente relevante al considerar los plazos extendidos asociados con la realización de estudios de IHQ y moleculares. Mientras que los resultados de una inmunohistoquímica pueden demorar hasta un mes desde la solicitud hasta la recepción, nuestro modelo proporciona predicciones prometedoras en un lapso temporal notablemente reducido.

Este factor adquiere suma importancia en el contexto de enfermedades críticas como el cáncer de endometrio, donde la prontitud en el diagnóstico se traduce directamente en la capacidad de seleccionar y aplicar tratamientos óptimos de manera oportuna. La rapidez y eficiencia de nuestro modelo no solo facilitan una intervención temprana, sino que también contribuyen significativamente a la optimización de los recursos médicos al reducir los tiempos de espera asociados con los métodos de diagnóstico tradicionales.

A pesar de que los resultados actuales del modelo han sido satisfactorios y muy prometedores, es importante reconocer que aún no alcanzan niveles óptimos y distan de poder ser implementados de manera efectiva en una institución médica. Si

bien exhiben una sensibilidad excelente, la especificidad y precisión no alcanzan los estándares requeridos para la tarea de clasificación en cuestión.

En primer lugar, las deficiencias del modelo se atribuyen principalmente a que la atención no se centra de manera exclusiva en las regiones tumorales. Como se evidencia en la Figura 5.8, entre los parches más relevantes para la clasificación se incluyen áreas de moco, hematíes y polimorfonucleares. La inestabilidad microsatelital es un fenómeno biológico que debe evaluarse de manera exclusiva en la región tumoral, lo cual representa una limitación significativa del modelo. Esta limitación conduce a la imposibilidad de elaborar conclusiones, a partir del modelo, sobre cuáles son las características tisulares distintivas de la inestabilidad microsatelital. Para abordar esta problemática, una solución viable consiste en incorporar mecanismos de atención focal para permitir que el modelo identifique la ubicación específica del tumor y se centre exclusivamente en ese tejido.

En particular, se podría explorar la alternativa propuesta de utilizar *vision transformers* para clasificar únicamente los parches tumorales de la imagen. Este enfoque resulta prometedor al trabajar exclusivamente con regiones tumorales, en vez de WSI, reduciendo significativamente el ruido en los datos de entrada.

Una limitación adicional del modelo radica en su entrenamiento con una cantidad limitada de muestras. Modelos destinados a analizar biomarcadores imperceptibles para el ojo humano requieren un conjunto extenso de datos para su formación. Aunque los autores de CLAM demostraron resultados satisfactorios al reducir su conjunto de datos a un número similar al nuestro, la diferencia crucial radica en que ellos se enfocan en clasificaciones de tumores, perceptibles en un microscopio de luz, por un patólogo entrenado. Contrariamente, esta investigación aborda una tarea mucho más compleja y ambiciosa, al analizar biomarcadores a nivel nuclear, fenómenos que ocurren a nivel molecular. Mejorar el rendimiento del modelo exige la obtención de un conjunto de muestras más extenso. No obstante, es relevante destacar que se utilizaron todos los conjuntos de datos públicos disponibles, los cuales abarcan la totalidad de los estudios publicados hasta la fecha. En este sentido, la limitación en la cantidad de muestras no se debe a una elección arbitraria, sino a la restricción impuesta por la disponibilidad de datos públicos. De todas formas, para mejorar el desempeño del modelo, una alternativa válida será conseguir nuevos especímenes etiquetados.

Finalmente, es importante considerar la diferencia en los métodos de clasificación

de las muestras entre los conjuntos de datos. Mientras que en el TCGA se empleó biología molecular, en el CPTAC y en el conjunto independiente se utilizó la evaluación del mecanismo MMR mediante IHQ como subrogante para la evaluación de la inestabilidad microsatelital. Aunque la discrepancia entre los métodos es mínima y se estima que no afecta significativamente el rendimiento del modelo, es necesario mencionarlo para una comprensión completa.

6.1. Comparación entre resultados dependientes e independientes

Como se ha detallado previamente en la sección correspondiente, se llevaron a cabo dos evaluaciones distintas: una empleando un subconjunto del conjunto de datos TCGA-CPTAC y otra utilizando un conjunto de imágenes independientes, obtenidas a partir de muestras retrospectivas de un laboratorio de anatomía patológica. Es importante señalar que las imágenes en el subconjunto de evaluación del TCGA-CPTAC no fueron todas adquiridas de la misma fuente ni digitalizadas de manera uniforme. De hecho, cada una de las fuentes, TCGA y CPTAC, representa una compilación de muestras provenientes de diversas instituciones. Incluso, las modalidades de digitalización varían entre TCGA y CPTAC. No obstante, a pesar de estas diferencias, las imágenes exhiben características similares que respaldan la validez de las evaluaciones realizadas. Adicionalmente, se destaca que, al contar con plena autonomía para seleccionar el umbral de decisión más apropiado para dicho conjunto, se logró alcanzar una sensibilidad del 100 %, acompañada por niveles aceptables de exactitud y precisión. Si bien la especificidad del modelo aún dista de los valores óptimos, estos valores sugieren que el modelo tiene un alto potencial para cumplir un rol fundamental en una institución médica.

Por otro lado, en lo que respecta al conjunto de muestras independientes, al mantener el umbral seleccionado, se observa una disminución en la satisfacción de las métricas, especialmente en la precisión y especificidad, las cuales aumentan las dudas respecto a su aceptabilidad. Aunque la sensibilidad aún se mantiene elevada, alcanzando un 91 %, las discrepancias más pronunciadas no radican exclusivamente en este aspecto, sino en la pérdida de precisión y especificidad del modelo.

Creemos que la razón más importante radica en la falta tanto de sistematización preanalítica como de generación de protocolos robustos. Estos últimos deben ser

controlados por las instituciones de salud, para que todas las técnicas de biología molecular que se efectúen en los tejidos sean evaluables y reproducibles. Debido a la falta de sistematización de procesos tanto preanalíticos como analíticos en los laboratorios en general, se generan enormes discrepancias que pueden asociarse a defectos técnicos como cambios de tinción, variaciones en el mantenimiento de las muestras en el laboratorio y disparidades en los procesos de digitalización que también contribuyen a la heterogeneidad evidente entre las imágenes. A pesar de los esfuerzos para mitigar estas diferencias mediante ajustes de parámetros durante el preprocesamiento, específicamente en la etapa de segmentación del tejido, o mediante la variación del umbral de decisión del modelo, las disparidades inherentes persisten y afectan el desempeño final del modelo.

Conocer la herramienta y sus necesidades técnicas, contribuirá a un mejor resultado global en los resultados.

Con el objetivo de ilustrar de manera más evidente las diferencias entre los conjuntos de datos, se adjuntan las Figuras 6.1 y 6.2, que exhiben una imagen representativa de cada conjunto, destacando de manera tangible las variaciones notables.

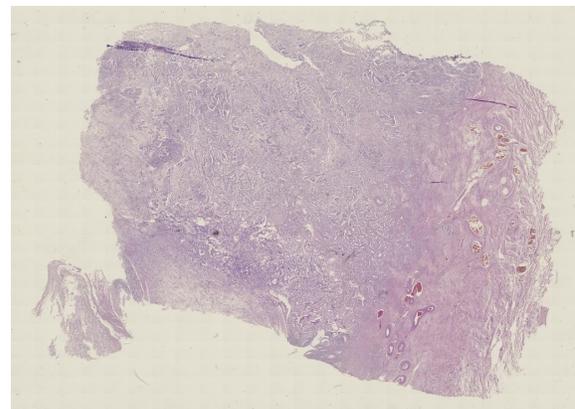


Figura 6.1: Imagen del conjunto TCGA-CPTAC

Figura 6.2: Imagen del conjunto independiente

Se puede observar la marcada diferencia en la tonalidad del fondo entre ambas imágenes: mientras que el fondo de la imagen del TCGA-CPTAC exhibe una tonalidad notablemente blanca, el fondo de la imagen del conjunto independiente presenta un matiz ligeramente amarillo. Asimismo, se evidencia una disparidad notable en la tinción de las muestras. Mientras que la imagen de la izquierda presenta un marca-

do contraste entre las coloraciones de la eosina y hematoxilina, la derecha presenta una tinción más descolorida y homogénea. Una imagen con mayor contraste entre sus tonalidades permite la identificación visual inmediata de distintas estructuras celulares, así como la distinción entre las tres capas del útero y la localización del tumor. Por otro lado, una imagen con menos contraste dificulta la interpretación del tejido no solo para el ojo humano sino que también para el algoritmo. Por esta razón es que se le adjudica la pérdida de desempeño del modelo a estas características de coloración que presentan las imágenes del conjunto independiente.

Al comparar los resultados del conjunto de evaluación TCGA-CPTAC con el independiente, queda en evidencia que si el objetivo es predecir la inestabilidad microsatelital en imágenes de archivo institucional, del laboratorio examinado o de otro con muestras similares, se deberán realizar ciertos cambios sobre los conjuntos de datos para mejorar el rendimiento del modelo. La opción óptima sería entrenar al modelo con imágenes de características similares a las del archivo retrospectivo con el que contamos, para que aprenda de imágenes cuyas condiciones no son óptimas y logre ver más allá del ruido, para darle mayor atención a las características importantes. Esto no es posible, dado que la cantidad de imágenes que se pudieron obtener no son suficientes como para formar un conjunto de entrenamiento. Es importante subrayar que, en la ejecución de este estudio, se recopilieron todas las muestras de cáncer de endometrio que fueron sometidas a inmunohistoquímica para la predicción de la inestabilidad microsatelital en el periodo comprendido entre los años 2019 y 2023. Cabe resaltar que previo a dicho período, este ensayo no era una práctica común y por lo tanto, la disponibilidad de muestras es limitada, ajustándose a aquellas obtenidas dentro del período de tiempo mencionado.

Como creemos que esta herramienta será utilizada con cotidianidad en el futuro en laboratorios y centros de salud, será mandatorio estandarizar lo mejor posible todos los procesos de los tejidos que deberán ser digitalizados y generar protocolos para el tratamiento que reciban las muestras, con el fin de optimizar la calidad de las mismas y disminuir las posibles alteraciones tisulares que dificultan la digitalización. Se deberá ajustar el grosor de los cortes histológicos, garantizar que los mismos no tengan pliegues significativos y ser meticulosos a la hora de la tinción con hematoxilina y eosina, para que las distintas estructuras se puedan ver con claridad. Será sumamente recomendable que las muestras sean digitalizadas de manera inmediata tras su obtención, garantizando condiciones óptimas para su preserva-

ción, con el objetivo de evitar cualquier posibilidad de deterioro. En términos de la digitalización, un aspecto crucial a considerar es la calidad del equipo utilizado para capturar las muestras. Es imperativo garantizar que dicho equipo sea capaz de capturar de manera precisa las características del tejido en cuestión. En conclusión, será importante contar con un protocolo sólido y estandarizado de tinción y fijación de muestras, junto con una digitalización inmediata llevada a cabo por un equipo de alta calidad.

6.2. Comparación con el estado del arte

Para poder discutir los resultados obtenidos en un contexto relevante, se realizó una búsqueda exhaustiva de trabajos relacionados. Cabe mencionar que este área de trabajo es considerablemente nueva, ya que se encontró únicamente una cantidad limitada de trabajos similares. A continuación, la Tabla 6.1 resume los resultados obtenidos tanto en este trabajo como en aquellos realizados por otros autores.

Tanto Hong et al [74] en su desarrollo del modelo *Panoptes*, como Fremont et al [75] en su desarrollo del modelo *im4MEC* buscan la clasificación de carcinomas endometriales en sus cuatro subtipos moleculares, entre ellos la inestabilidad microsatelital.

El modelo *Panoptes* está basado en la red de aprendizaje profundo *Inception ResNet-V1* y fue entrenado con un conjunto de 496 imágenes obtenidas de las fuentes TCGA-UCEC y CPTAC-UCEC, al igual que en el presente trabajo. Tuvieron acceso a toda la información genómica de estas muestras, lo cual explica por qué su grupo de imágenes es mayor. Al igual que en nuestro estudio, también buscaron evaluar el modelo en un conjunto de imágenes independiente, que consiste de 137 imágenes de la Universidad de Nueva York (NYU).

Por otro lado, el modelo *im4MEC* consta de una arquitectura que incluye dos modelos, uno de auto-supervisión para la extracción de características y uno de atención para la clasificación de tumores de endometrio en uno de los cuatro subtipos moleculares. Fue entrenado con un conjunto de 1.635 imágenes, obtenidas tanto del TCGA como de bases de datos no abiertas al público. El modelo fue evaluado en un conjunto independiente de 393 imágenes del ensayo clínico PORTEC-3.

Por último, el trabajo realizado por Zhang et al [84] utiliza redes neuronales para la creación de un modelo que realiza la predicción de inestabilidad microsatelital en

cáncer de endometrio, llamado GAM-VGG11. La base de datos que se utiliza es también el proyecto TCGA-UCEC, del cual se toman 70 imágenes para el entrenamiento y 25 para la evaluación. Los resultados de los tres modelos para la clasificación de tumores según su estado microsatelital se resumen en la siguiente tabla.

Dataset	Modelo	Magnificación	AUROC	Exactitud	Precisión	Sensibilidad	Especificidad	NPV
Dependiente	Panoptes	10x	0,83	0,78	0,73	0,53	0,90	0,79
	GAM-VGG11	20x	0,80	0,80	0,80	0,86	0,73	0,80
	Modelo_balanceado_10x	10x	0,92	0,82	0,70	1,00	0,68	1,00
Independiente	Panoptes	10x	0,66	0,20	0,50	0,04	0,83	0,18
	im4MEC	40x	0,84	0,76	0,71	0,74	0,73	0,88
	Modelo_balanceado_10x	10x	0,74	0,63	0,54	0,91	0,41	0,86

Tabla 6.1: Métricas estadísticas de los modelos *Panoptes*, *GAM-VGG11*, *im4MEC* y *Modelo_balanceado_10x*

La sección superior de la tabla presenta los resultados correspondientes a una evaluación llevada a cabo utilizando un subconjunto de datos de evaluación proveniente del mismo conjunto de datos utilizado en el proceso de entrenamiento. En contraste, la sección inferior muestra los resultados de evaluaciones realizadas con un conjunto de datos independiente, es decir, imágenes que tienen una fuente distinta a aquella empleada para entrenar al modelo.

Analizando los resultados dependientes, se puede observar que, de los tres modelos comparados, el *modelo_balanceado_10x* obtuvo los mejores resultados para todas las métricas salvo la precisión y especificidad. Al seleccionar un umbral que le da prioridad a la sensibilidad por sobre las otras métricas, se puede explicar por qué los resultados obtenidos para la precisión y especificidad no son los óptimos. De todas formas, como fue explicado anteriormente, los falsos positivos no confieren una gran preocupación dado que se van a poder detectar más adelante en el laboratorio. Al realizar una evaluación exhaustiva de todas las métricas disponibles, es evidente que nuestro modelo de clasificación ha demostrado resultados consistentes, sino superiores en varios aspectos a los modelos existentes en el campo. Cabe destacar que la novedad de nuestro enfoque es particularmente significativa, dado que la literatura existente en esta área presenta escasez de estudios y resultados poco robustos. Esta falta de investigaciones sólidas resalta aún más el valor de nuestro estudio, ya que hemos contribuido con nuevas perspectivas y la consistencia y superioridad de nuestros resultados subrayan la importancia de explorar en profundidad este nicho poco

explorado, proporcionando una base más sólida para futuras investigaciones en este campo emergente.

Siguiendo con la parte inferior de la tabla, se pueden observar las métricas de los distintos modelos al ser evaluados en un conjunto independiente. La sensibilidad fue superior para nuestro modelo, mientras que *im4MEC* obtuvo los mejores resultados para las otras métricas y *Panoptes* tuvo el peor desempeño global. El modelo *im4MEC* es evaluado en un conjunto que corresponde a un ensayo clínico realizado por el *Gynecological Cancer Group*. Se presume que la calidad de las muestras y su digitalización, a diferencia de las imágenes en nuestro conjunto independiente, es superior y se asemeja más a las del TCGA-CPTAC. Esto concuerda con las conclusiones derivadas en la sección previa, las cuales indican que las etapas subsiguientes para mejorar el desempeño del modelo implican el perfeccionamiento del conjunto independiente. Además dicho estudio cuenta con el conjunto de entrenamiento más grande de todos, se entrena el modelo con aproximadamente cuatro veces más imágenes que el nuestro. Nuevamente, esto apoya la hipótesis de que aumentando la cantidad de muestras con las que se entrena al modelo, se podría mejorar drásticamente el desempeño del mismo.

6.3. Futuras direcciones

Este estudio ha demostrado la viabilidad y el potencial de emplear un modelo de aprendizaje profundo para la clasificación de tumores de endometrio basándose en su estado de inestabilidad microsatelital. A partir de los resultados obtenidos, se propone la hipótesis de que la misma arquitectura, o una similar, podría aplicarse con éxito para la clasificación de tumores según la presencia de otros biomarcadores, como el gen Pol-E y la proteína p53, utilizados en la clasificación molecular de los tumores de endometrio. Esta clasificación resulta esencial para determinar el pronóstico de la paciente y establecer el tratamiento más adecuado. No sólo se podría expandir el abordaje de esta arquitectura sino que se debería continuar con la exploración de otras líneas de trabajo para obtener resultados aún más satisfactorios.

Adicionalmente, la versatilidad de esta herramienta se proyecta hacia la capacidad de procesar imágenes de distintos tipos de tumores, ya sea para predecir el estado microsatelital o detectar otros biomarcadores en diversos órganos. Al demostrar que un modelo, mediante el uso de imágenes de tinciones H&E, puede clasificar con

exactitud los tumores de endometrio en función de un biomarcador específico, se abre la posibilidad de extender la aplicación del aprendizaje profundo a una variedad de tareas. La detección de múltiples biomarcadores a partir de una única imagen refuerza la justificación para llevar a cabo la digitalización adecuada de las muestras, incluso respaldando la inversión en un digitalizador. Con una única imagen y una interfaz que integre varios modelos, es factible obtener resultados que, en el ámbito de un laboratorio, requiere múltiples y prolongados ensayos.

Conclusión

En conclusión, a través de este estudio, queda demostrada la viabilidad del uso de inteligencia artificial para la clasificación de tumores de endometrio según su estado de inestabilidad microsatelital. Mediante el uso de la arquitectura CLAM, fue posible la creación de un modelo de aprendizaje profundo que lleva a cabo esta tarea de forma efectiva, buscando superar las limitaciones de las estrategias tradicionales empleadas en el laboratorio. Adicionalmente, se desarrolla una interfaz gráfica para facilitar el uso del modelo, de tal forma que sea accesible para cualquier profesional de la salud dentro de un centro médico.

En oncología, el tiempo es vida. Es por esto que es de gran relevancia el desarrollo de una herramienta como la propuesta en este trabajo, que pueda acortar los tiempos de diagnóstico de manera significativa, para que el paciente pueda recibir un tratamiento personalizado de forma eficaz. A medida que se proyecta el futuro de esta herramienta, se considera su versatilidad en el procesamiento de imágenes de diferentes tipos de tumores, ofreciendo la capacidad de predecir el estado microsatelital o detectar otros biomarcadores en diversos órganos. Expandir su uso para la clasificación de tumores según la presencia del gen Pol-E y la proteína p53, sería fundamental para la determinación del pronóstico y la elección del tratamiento adecuado para una paciente. La aplicación exitosa de este modelo a través de imágenes de tinciones H&E para la clasificación precisa de tumores, respalda la necesidad de una digitalización adecuada de las muestras, e incluso justifica la inversión en tecnologías de digitalización.

Este trabajo deja abierta la posibilidad de que, con el desarrollo adecuado, se pueda implementar una interfaz como esta para proporcionar resultados, que en entornos de laboratorio requerirían múltiples y prolongados ensayos. Aunque se reconoce que los resultados actuales pueden no alcanzar niveles óptimos y aún queda mucho por hacer en cuanto a mejorar el desempeño del modelo, se abre la puerta a futuras investigaciones y refinamientos para perfeccionar aún más este enfoque innovador.

En términos prácticos, se subraya la importancia de considerar cuidadosamente las condiciones de obtención y digitalización de las muestras, así como la relevancia de optimizar parámetros clave durante el procesamiento para garantizar la coherencia y validez de los resultados. Queda en evidencia que la patología digital es un paso hacia el futuro de la medicina, haciendo posible el uso de una gran cantidad de herramientas. A pesar de las limitaciones actuales, este trabajo proporciona una base sólida y prometedora para el uso futuro de modelos de inteligencia artificial en el ámbito de la patología molecular y la clasificación precisa de tumores, abriendo nuevas posibilidades para la mejora continua en el diagnóstico y tratamiento del cáncer de endometrio y otras neoplasias.

Abreviaturas

ADN: Ácido desoxirribonucleico

ASR: *Average Standard Rate*

AUROC: Área debajo de la curva ROC (Ref. ROC debajo)

AUC-PR: Área debajo de la curva PR (Ref. PR debajo)

BWH: *Brigham and Women's Hospital*

CI: intervalo de confianza

CLAM: *Clustering Constrained Attention Multiple Instance Learning*

CNN: Red neuronal convolucional

CPTAC: *Clinical Proteomic Tumor Analysis Consortium*

CPU: *Computer Processing Unit*

FDA: Administración de Alimentos y Medicamentos de Estados Unidos

FIGO: Federación internacional de ginecología y obstetricia

FN: Falso negativo

FP: Falso positivo

GB: *giga bytes*

GPU: *Graphics Processing Unit*

HSV: *Hue Saturation Value*

H&E: Hematoxilina y Eosina

IA: Inteligencia artificial

IHQ: Inmunohistoquímica

ITBA: Instituto Tecnológico de Buenos Aires

JPG: *Joint Photographic Experts Group*

MIL: *Múltiples Instance Learning*

ML: *Machine learning*

MMR: *DNA Mismatch repair*

MMRd: *deficient MMR*

MMRp: *proficient MMR*

mpp: Micrones por pixel

MSI: Inestabilidad microsatelital
MSI-H: Inestabilidad microsatelital *high*
MSI-L: Inestabilidad microsatelital *low*
MSS: Estabilidad microsatelital
NCI: Instituto nacional de cáncer
NPV: Valor predictivo negativo
NSCLC: Cáncer de pulmón de células no pequeñas
NSMP: *Non-Specific Molecular Profile*
NYU: Universidad de Nueva York
OMS: Organización mundial de la salud
PCR: Reacción en cadena de la polimerasa
PET-CT: Tomografía computarizada por emisión de positrones
PNG: Portable Network Graphics
PPV: Valor predictivo positivo
PR: Precisión - recall
RAM: *Random-access Memory*
RCC: Carcinoma de célula renal
RGB: *Red Green Blue*
ROC: característica operativa del receptor
ROI: *Region of interest*
SVM: *Support Vector Machines*
SVS: *Slide Viewer System*
TCGA: *The Cancer Genome Atlas*
TNM: *Tumor node metastasis*
UCEC: *Uterine Corpus Endometrial Carcinoma*
ViT: *Vision Transformers*
VN: Verdadero negativo
VP: Verdadero positivo
VRAM: RAM virtual
WSI: *Whole Slide Images*

Anexo

9.1. Clasificación histológica del cáncer de cuerpo de útero según la OMS

WHO classification of tumours of the uterine corpus

Endometrial epithelial tumours and precursors	8897/1	Smooth muscle tumour of uncertain malignant potential
Endometrial hyperplasia without atypia	8891/1*	Epithelioid smooth muscle tumour of uncertain malignant potential
8380/2 Atypical hyperplasia of the endometrium	8896/1*	Myxoid smooth muscle tumour of uncertain malignant potential
8380/3 Endometrioid adenocarcinoma NOS		Spindle smooth muscle tumour of uncertain malignant potential
POLE-ultramutated endometrioid carcinoma	8898/1	Metastasizing leiomyoma
Mismatch repair-deficient endometrioid carcinoma	8890/3	Leiomyosarcoma NOS
p53-mutant endometrioid carcinoma		Spindle leiomyosarcoma
No specific molecular profile (NSMP) endometrioid carcinoma	8891/3	Epithelioid leiomyosarcoma
8441/3 Serous carcinoma NOS	8896/3	Myxoid leiomyosarcoma
8310/3 Clear cell adenocarcinoma NOS	8930/0	Endometrial stromal nodule
8020/3 Carcinoma, undifferentiated, NOS	8931/3	Endometrial stromal sarcoma, low grade
8323/3 Mixed cell adenocarcinoma	8930/3	Endometrial stromal sarcoma, high grade
9110/3 Mesonephric adenocarcinoma	8805/3	Undifferentiated sarcoma
8070/3 Squamous cell carcinoma NOS	8590/1	Uterine tumour resembling ovarian sex cord tumour
8144/3 Mucinous carcinoma, intestinal type	8714/0	Perivascular epithelioid tumour, benign
9111/3* Mesonephric-like adenocarcinoma	8714/3	Perivascular epithelioid tumour, malignant
8980/3 Carcinosarcoma NOS	8825/1	Inflammatory myofibroblastic tumour
		Epithelioid myofibroblastic sarcoma
Tumour-like lesions		
Endometrial polyp		
Endometrial metaplasia		
Arias-Stella reaction		
Mesenchymal tumours specific to the uterus		Mixed epithelial and mesenchymal tumours
8890/0 Leiomyoma NOS	8932/0	Adenomyoma NOS
8890/0 Lipoleiomyoma	8932/0	Atypical polypoid adenomyoma
8890/0 Leiomyoma, apoplectic	8933/3	Adenosarcoma
8890/0 Leiomyoma, hydropic		
8890/0 Dissecting leiomyoma		Miscellaneous tumours
8892/0 Cellular leiomyoma	9473/3	Primitive neuroectodermal tumour NOS
8896/0 Myxoid leiomyoma	9064/3	Germ cell tumour NOS
8891/0 Epithelioid leiomyoma	9071/3	Yolk sac tumour NOS
8893/0 Symplastic leiomyoma	9080/0	Mature teratoma NOS
8890/1 Leiomyomatosis NOS	9080/3	Immature teratoma NOS
8890/1 Intravenous leiomyomatosis		

Figura 9.1: Clasificación histológica de la OMS para el cáncer de endometrio [85]

9.2. Parámetros para la segmentación

Procedencia de la imagen	sthresh	mthresh	close	use_otsu
TCGA	8	7	4	false
CPTAC	15	11	2	false
Independiente biopsia	15	11	2	true
Independiente resección	8	7	4	true

Tabla 9.1: Parámetros para la segmentación del tejido para cada imagen

Donde:

- sthresh representa el umbral para la saturación
- mthresh representa el tamaño del filtro de mediana
- close representa el tamaño del elemento estructurante para el cierre morfológico
- use_otsu si se utiliza este método para la binarización de la imagen

9.3. Cálculo del puntaje de atención para cada parche

$$a_{i,k} = \frac{\exp\{W_{a,i}(\tanh(V_a \cdot h_k) \odot \text{sigm}(U_a \cdot h_k))\}}{\sum_{j=1}^K \exp\{W_{a,i}(\tanh(V_a \cdot h_j) \odot \text{sigm}(U_a \cdot h_j))\}} \quad (9.1)$$

Donde:

- $a_{i,k}$ es el puntaje de atención para la clase i del parche k
- $W_{a,i}$ es la primera capa de la rama de atención para la clase i
- V_a y U_a son capas compartidas entre clases, es decir, parte del núcleo de atención
- h_k es el vector de 512 elementos del parche k , obtenido luego de comprimir el vector de 1024 características

- K es la totalidad de parches de la imagen

9.4. Optimizador Adam

La fórmula para el optimizador Adam es la siguiente [86]:

$$w_{t+1} = w_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon} - \lambda \cdot w_t \quad (9.2)$$

Donde:

- w_{t+1} son los pesos actualizados
- w_t son los pesos actuales
- η es la tasa de aprendizaje o *learning rate*, 2×10^{-4}
- m_t es el primer momento (media móvil de los gradientes)
- v_t es el segundo momento (media móvil de los cuadrados de los gradientes)
- ϵ es un término para evitar la división por cero, 1×10^{-8}
- λ es la tasa de decaimiento del peso, 1×10^{-5}

Las fórmulas utilizadas para calcular los momentos del optimizador son las siguientes:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (9.3)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (9.4)$$

Donde:

- β_1 es el parámetro de suavizado exponencial para el primer momento, 0,9
- β_2 es el parámetro de suavizado exponencial para el segundo momento, 0,999
- m_{t-1} y v_{t-1} son las estimaciones anteriores de los momentos
- g_t es el gradiente en el paso del tiempo t

Durante el proceso de entrenamiento, se utiliza también un *drop-out* de 0,25 luego de cada capa del núcleo de atención para la regularización del modelo.

9.5. Matrices de confusión y curvas PR antes de la umbralización

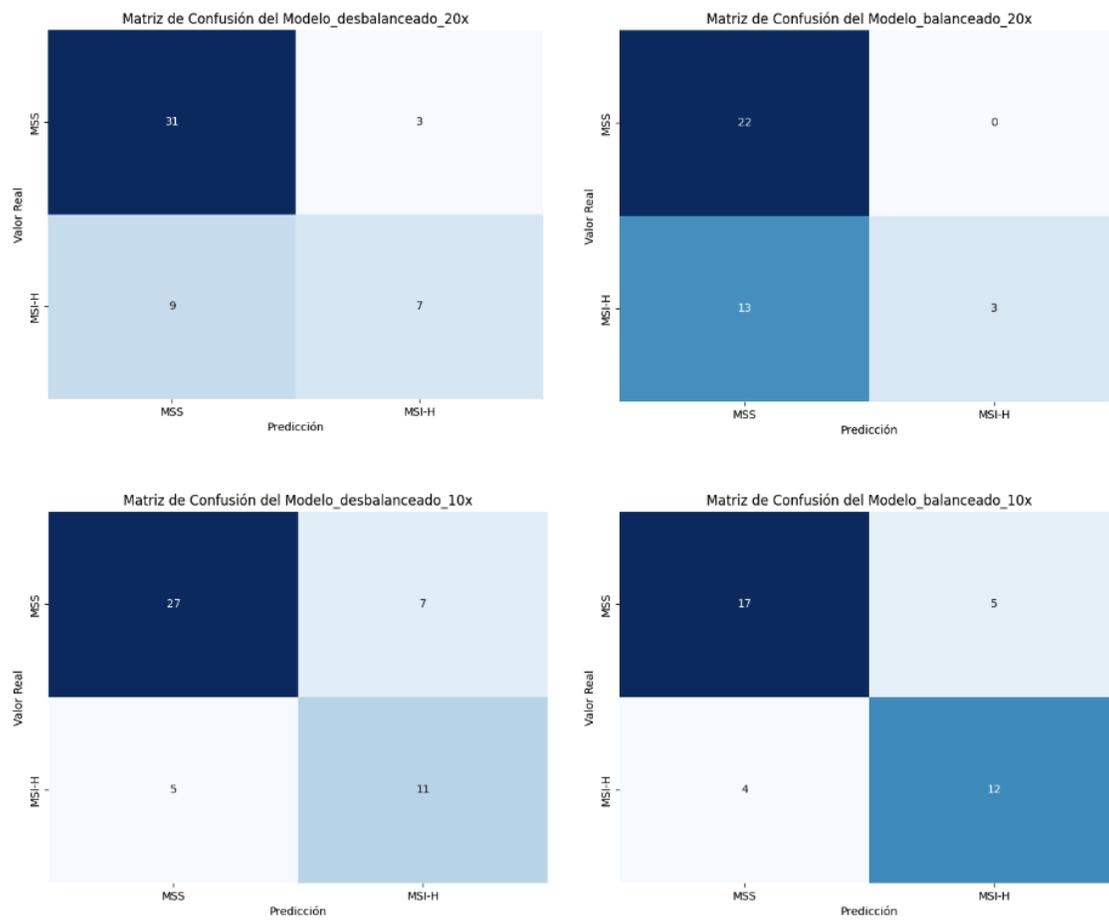


Figura 9.2: Matrices de confusión para los cuatro modelos antes de la umbralización óptima

Tabla de métricas de la curva PR:

	Umbral PR	Precision	Recall	AUC-PR
0	0.053788	0.320000	1.0000	0.70537
1	0.111780	0.340426	1.0000	0.70537
2	0.128494	0.363636	1.0000	0.70537
3	0.156292	0.365854	0.9375	0.70537
4	0.180563	0.394737	0.9375	0.70537
5	0.205516	0.428571	0.9375	0.70537
6	0.215349	0.468750	0.9375	0.70537
7	0.264162	0.517241	0.9375	0.70537
8	0.277466	0.500000	0.8125	0.70537
9	0.314939	0.478261	0.6875	0.70537
10	0.350627	0.500000	0.6250	0.70537
11	0.417687	0.529412	0.5625	0.70537
12	0.438817	0.571429	0.5000	0.70537
13	0.499101	0.636364	0.4375	0.70537
14	0.548147	0.750000	0.3750	0.70537
15	0.585799	1.000000	0.3125	0.70537
16	0.615163	1.000000	0.1250	0.70537

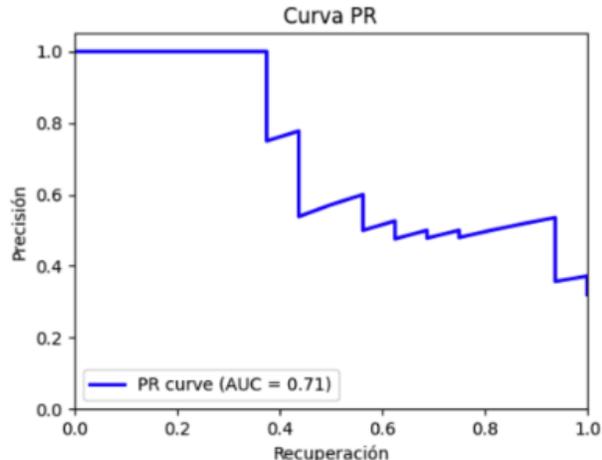


Figura 9.3: Curva PR y posibles umbrales para el modelo_desbalanceado_20x

Tabla de métricas de la curva PR:

	Umbral PR	Precision	Recall	AUC-PR
0	0.061460	0.320000	1.0000	0.588597
1	0.121690	0.340426	1.0000	0.588597
2	0.160193	0.363636	1.0000	0.588597
3	0.178396	0.390244	1.0000	0.588597
4	0.250628	0.421053	1.0000	0.588597
5	0.317376	0.428571	0.9375	0.588597
6	0.366897	0.437500	0.8750	0.588597
7	0.382546	0.482759	0.8750	0.588597
8	0.426890	0.461538	0.7500	0.588597
9	0.451421	0.521739	0.7500	0.588597
10	0.489011	0.600000	0.7500	0.588597
11	0.513499	0.647059	0.6875	0.588597
12	0.540530	0.642857	0.5625	0.588597
13	0.590885	0.727273	0.5000	0.588597
14	0.726219	0.625000	0.3125	0.588597
15	0.762053	0.800000	0.2500	0.588597
16	0.855939	0.500000	0.0625	0.588597

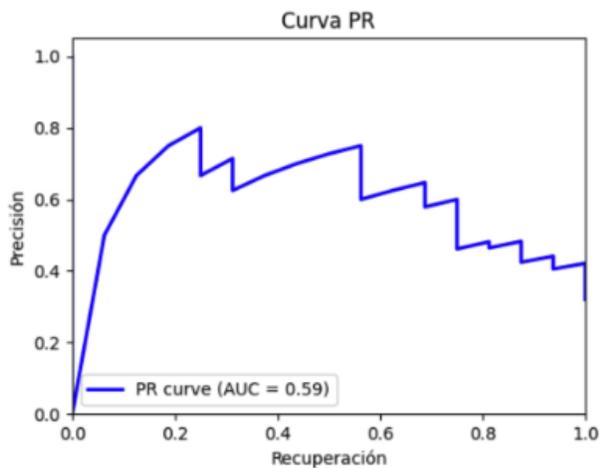


Figura 9.4: Curva PR y posibles umbrales para el modelo_desbalanceado_10x

Tabla de métricas de la curva PR:

	Umbral PR	Precision	Recall	AUC-PR
0	0.130023	0.421053	1.0000	0.803446
1	0.186607	0.457143	1.0000	0.803446
2	0.243684	0.468750	0.9375	0.803446
3	0.272459	0.482759	0.8750	0.803446
4	0.307268	0.538462	0.8750	0.803446
5	0.344304	0.608696	0.8750	0.803446
6	0.367504	0.650000	0.8125	0.803446
7	0.381254	0.705882	0.7500	0.803446
8	0.407572	0.785714	0.6875	0.803446
9	0.449565	0.818182	0.5625	0.803446
10	0.469845	0.750000	0.3750	0.803446
11	0.491650	1.000000	0.3125	0.803446
12	0.579999	1.000000	0.1250	0.803446

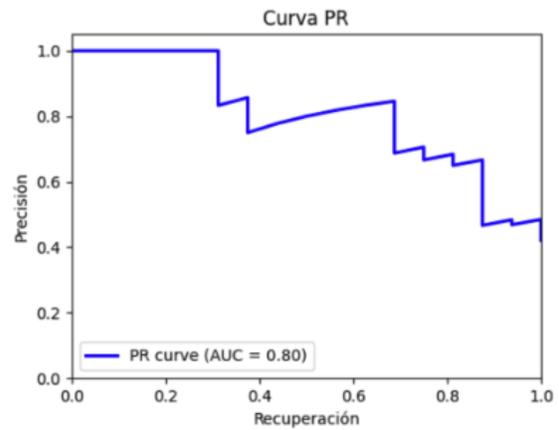


Figura 9.5: Curva PR y posibles umbrales para el modelo_balanceado_20x

Referencias

[1] About Digital Pathology (no date) DPA. Available at: <https://digitalpathologyassociation.org/about-digital-pathology> (Accessed: 30 October 2023).

[2] The International Agency for Research on Cancer (IARC) (no date) Global cancer observatory, Global Cancer Observatory. Available at: <https://gco.iarc.fr/> (Accessed: 23 October 2023).

[3] Real, B. del (2023) Cáncer de endometrio-útero, SEOM. Available at: <https://seom.org/171-Informaci%C3%B3n%20al%20P%C3%BAblico%20-%20Patolog%C3%ADas/cancer-de-endometrio-utero#:~:text=Un%2075%2D80%25%20de%20los,5%20aos%20de%20un%2095%25>. (Accessed: 23 October 2023).

[4] “What is endometrial cancer?: Uterine cancer,” Uterine Cancer — American Cancer Society, <https://www.cancer.org/cancer/types/endometrial-cancer/about/what-is-endometrial-cancer.html> (accessed Sep. 28, 2023).

[5] C. Lazzarino, “Fisiología de la Reproducción,” in Fisiología , Sep. 28, 2023

[6] S. Lee, “The uterus,” Canadian Cancer Society, <https://cancer.ca/en/cancer-information/cancer-types/uterine/what-is-uterine-cancer/the-uterus#:~:text=The%20perimetrium%20is%20the%20outer,the%20organs%20of%20the%20pelvis>. (accessed Sep. 28, 2023).

[7] About Digital Pathology (no date) DPA. Available at: <https://digitalpathologyassociation.org/about-digital-pathology> (Accessed: 30 October 2023).

[8] professional, C.C. medical (no date) Uterine (endometrial) cancer: What is it?, Cleveland Clinic. Available at: <https://my.clevelandclinic.org/health/diseases/16409-uterine-cancer> (Accessed: 31 October 2023)

[9] The International Agency for Research on Cancer (IARC) (no date) Global cancer observatory, Global Cancer Observatory. Available at: <https://gco.iarc.fr/> (Accessed: 23 October 2023).

[10] Estadísticas - Incidencia (2021) Argentina.gob.ar. Available at: <https://www.argentina.gob.ar/salud/instituto-nacional-del-cancer/estadisticas/incidencia> (Accessed: 23 October 2023).

[11] Key statistics for endometrial cancer (no date) American Cancer Society. Available at: <https://www.cancer.org/cancer/types/endometrial-cancer/about/key-statistics.html> (Accessed: 23 October 2023).

[12] N. A. J. Ryan et al., “The proportion of endometrial cancers associated with Lynch Syndrome: A systematic review of the literature and meta-analysis,” *Genetics in Medicine*, vol. 21, no. 10, pp. 2167–2180, 2019. doi:10.1038/s41436-019-0536-8

[13] Tratamiento del Cáncer de Endometrio (PDQ®) (2023) Instituto Nacional del Cáncer. Available at: <https://www.cancer.gov/espanol/tipos/uterino/pro/tratamiento-endometrio-pdq> (Accessed: 25 October 2023).

[14] D. de Biase *et al.*, “Integrated Clinicopathologic and Molecular Analysis of endometrial carcinoma: Prognostic impact of the new ESGO-ESTRO-ESP Endometrial Cancer Risk Classification and proposal of histopathologic algorithm for its implementation in clinical practice,” *Frontiers in Medicine*, vol. 10, 2023. doi:10.3389/fmed.2023.1146499

[15] “Microsatellite,” Genome.gov, <https://www.genome.gov/genetics-glossary/Microsatellite> (accessed Sep. 28, 2023).

[16] K. A. Feldheim, “Microsatellite,” *Brenner’s Encyclopedia of Genetics*, pp. 412–413, 2013. doi:10.1016/b978-0-12-374984-0.00947-5

-
- [17] J. K. Sehn, “Insertions and deletions (indels),” *Clinical Genomics*, pp. 129–150, 2015. doi:10.1016/b978-0-12-404748-8.00009-5
- [18] G. Tallen, Y. Yang, and K. Riabowol, “Epigenetic regulation of the cell cycle & DNA-repair in cancer,” *Genome Stability*, pp. 521–531, 2021. doi:10.1016/b978-0-323-85679-9.00027-1
- [19] P. Hsieh, “DNA mismatch repair in disease and aging,” *Encyclopedia of Biological Chemistry*, pp. 60–65, 2013. doi:10.1016/b978-0-12-378630-2.00304-2
- [20] W. Wu, Y. Liu, S. Zeng, Y. Han, and H. Shen, “Intratumor heterogeneity: The Hidden Barrier to immunotherapy against MSI tumors from the perspective of IFN- γ signaling and tumor-infiltrating lymphocytes,” *Journal of Hematology & Oncology*, vol. 14, no. 1, 2021. doi:10.1186/s13045-021-01166-3
- [21] C. Lefol et al., “Acquired somatic MMR deficiency is a major cause of MSI tumor in patients suspected for ‘Lynch-like syndrome’ including young patients,” *European Journal of Human Genetics*, vol. 29, no. 3, pp. 482–488, 2020. doi:10.1038/s41431-020-00778-6
- [22] “Metilación,” *Genome.gov*, <https://www.genome.gov/es/genetics-glossary/Methylation> (accessed Oct. 31, 2023).
- [23] S. Magaki, S. A. Hojat, B. Wei, A. So, and W. H. Yong, “An introduction to the performance of Immunohistochemistry,” *Methods in Molecular Biology*, pp. 289–298, 2018. doi:10.1007/978-1-4939-8935-5_25
- [24] R. Parra-Medina and J. F. Polo, “Inmunofluorescencia en Tejidos Fijados y preservados en parafina (if-P). Una mirada desde la Patología Quirúrgica,” *Repertorio de Medicina y Cirugía*, vol. 26, no. 4, pp. 202–207, 2017. doi:10.1016/j.reper.2017.08.009

- [25] “Ventana MMR IHC panel - Roche Diagnostics USA,” Roche Diagnostics, <https://diagnostics.roche.com/content/dam/diagnostics/us/en/products/c/colorectal-ihc-portfolio/MMR-IHC-Panel-InterpretationGuide.pdf> (accessed Oct. 24, 2023).
- [26] “Immunohistochemistry: Antibody staining: IHC detection,” Antibody Staining — IHC Detection, <https://www.bdbiosciences.com/en-us/learn/applications/immunohistochemistry#0verview> (accessed Nov. 11, 2023).
- [27] “Reacción en Cadena de la polimerasa (PCR) (artículo),” Khan Academy, <https://es.khanacademy.org/science/ap-biology/gene-expression-and-regulation/biotechnology/a/polymerase-chain-reaction-pcr> (accessed Oct. 25, 2023).
- [28] N. Suraweera et al., “Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and PENTAPLEX PCR,” *Gastroenterology*, vol. 123, no. 6, pp. 1804–1811, 2002. doi:10.1053/gast.2002.37070
- [29] K. C. Kurnit, S. N. Westin, and R. L. Coleman, “Microsatellite instability in endometrial cancer: New purpose for an old test,” *Cancer*, vol. 125, no. 13, pp. 2154–2163, 2019. doi:10.1002/ncr.32058
- [30] L. Wang et al., “Hot and cold tumors: Immunological features and the therapeutic strategies,” *MedComm*, vol. 4, no. 5, 2023. doi:10.1002/mco2.343
- [31] X. Ren et al., “Immunological classification of tumor types and advances in precision combination immunotherapy,” *Frontiers*, <https://www.frontiersin.org/articles/10.3389/fimmu.2022.790113/full> (accessed Sep. 29, 2023).
- [32] J. M. Lancaster, C. B. Powell, L. Chen, and D. L. Richardson, “Society of Gynecologic Oncology Statement on risk assessment for inherited gynecologic cancer predispositions,” *Gynecologic Oncology*, vol. 136, no. 1, pp. 3–7, 2015. doi:10.1016/j.ygyno.2014.09.009

- [33] J. Galon and D. Bruni, “Approaches to treat immune hot, altered and cold tumours with combination immunotherapies,” *Nature Reviews Drug Discovery*, vol. 18, no. 3, pp. 197–218, 2019. doi:10.1038/s41573-018-0007-y
- [34] Quimioterapia para Tratar el cáncer (2015) Instituto Nacional del Cáncer. Available at: <https://www.cancer.gov/espanol/cancer/tratamiento/tipos/quimioterapia> (Accessed: 25 October 2023).
- [35] “Pembrolizumab,” National Cancer Institute, <https://www.cancer.gov/about-cancer/treatment/drugs/pembrolizumab> (accessed Oct. 24, 2023)
- [36] “What is Artificial Intelligence (AI)?,” IBM, <https://www.ibm.com/topics/artificial-intelligence> (accessed Oct. 30, 2023).
- [37] “Ai index report 2023 – artificial intelligence index,” Stanford University — AI Index, <https://aiindex.stanford.edu/report/> (accessed Nov. 11, 2023).
- [38] “What is machine learning?,” IBM, <https://www.ibm.com/topics/machine-learning>(accessed Oct. 30, 2023).
- [39] “What is deep learning?,” IBM, <https://www.ibm.com/topics/deep-learning#:~:text=the%20next%20step-,What%20is%20deep%20learning%3F,from%20large%20amounts%20of%20data.> (accessed Oct. 30, 2023).
- [40] P. Huet, “Qué son las redes neuronales y sus aplicaciones,” OpenWebinars.net, <https://openwebinars.net/blog/que-son-las-redes-neuronales-y-sus-aplicaciones/> (accessed Oct. 30, 2023)
- [41] “Convolutional layer,” Convolutional Layer - an overview — ScienceDirect Topics, <https://www.sciencedirect.com/topics/engineering/convolutional-layer#:~:text=A%20convolutional%20layer%20is%20the,and%20creates%20an%20activation%20map.> (accessed Nov. 2, 2023).

- [42] “Convolutional Neural Network,” Convolutional Neural Network - an overview — ScienceDirect Topics, <https://www.sciencedirect.com/topics/mathematics/convolutional-neural-network> (accessed Nov. 2, 2023).
- [43] “Attention Layer,” Deep Learning for molecules and materials, <https://dmol.pub/dl/attention.html#:~:text=Attention%20layers%20are%20inspired%20by,and%20the%20values%20are%20equal> (accessed Nov. 11, 2023).
- [44] X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, “An empirical study of spatial attention mechanisms in Deep Networks,” 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019. doi:10.1109/iccv.2019.00679
- [45] J. Brownlee, “A gentle introduction to dropout for Regularizing Deep Neural Networks,” MachineLearningMastery.com, <https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/> (accessed Nov. 2, 2023).
- [46] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, “Impact of fully connected layers on performance of convolutional neural networks for Image Classification,” Neurocomputing, vol. 378, pp. 112–119, 2020. doi:10.1016/j.neucom.2019.10.008
- [47] K. E. Koech, “Softmax activation function-how it actually works,” Medium, <https://towardsdatascience.com/softmax-activation-function-how-it-actually-works-d292d335bd78> (accessed Nov. 2, 2023).
- [48] Swapna., “Convolutional Neural Network: Deep learning,” Developers Breach, <https://developersbreach.com/convolution-neural-network-deep-learning/> (accessed Oct. 30, 2023).
- [49] “Inicialización de los Parámetros,” Inicialización de los parámetros — Interactive Chaos, <https://interactivechaos.com/es/manual/tutorial-de-deep-learning/inicializacion-de-los-parametros> (accessed Nov. 2, 2023).

- [50] R. Alberto, “Redes Neuronales: Propagación Hacia Adelante y propagación hacia atrás,” Medium, <https://rubialesalberto.medium.com/redes-neuronales-propagaci%C3%B3n-hacia-adelante-y-propagaci%C3%B3n-hacia-atr%C3%A1s-4745c0fb6286> (accessed Nov. 11, 2023).
- [51] J. Brownlee, “A gentle introduction to cross-entropy for Machine Learning,” MachineLearningMastery.com, <https://machinelearningmastery.com/cross-entropy-for-machine-learning/> (accessed Nov. 2, 2023).
- [52] R. Balasubramanian, “Loss functions-multiclass SVM loss and cross entropy loss,” Medium, <https://medium.com/analytics-vidhya/loss-functions-multiclass-svm-loss-and-cross-entropy-loss-9190c68f13e0> (accessed Nov. 2, 2023).
- [53] J. Luo, H. Qiao, and B. Zhang, “Learning with Smooth Hinge losses,” *Neurocomputing*, vol. 463, pp. 379–387, 2021. doi:10.1016/j.neucom.2021.08.060
- [54] J. Brownlee, “Gentle introduction to the adam optimization algorithm for deep learning,” MachineLearningMastery.com, <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/> (accessed Nov. 2, 2023).
- [55] J. Delua, “Supervised vs. unsupervised learning: What’s the difference?,” IBM Blog, <https://www.ibm.com/blog/supervised-vs-unsupervised-learning/> (accessed Oct. 30, 2023).
- [56] G. Campanella 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. doi:10.1038/s41591-019-0508-1
- [57] P. Maia, “An introduction to multiple instance learning,” NILG.AI, <https://nilg.ai/202105/an-introduction-to-multiple-instance-learning/> (accessed Oct. 30, 2023).

- [58] J. Boschman, “Multiple-instance learning: One minute introduction,” Medium, <https://medium.com/one-minute-machine-learning/multiple-instance-learning-one-minute-introduction-c1a382936105> (accessed Nov. 14, 2023).
- [59] “Transfer learning and fine-tuning; Tensorflow Core,” TensorFlow, https://www.tensorflow.org/tutorials/images/transfer_learning (accessed Oct. 31, 2023).
- [60] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” arXiv.org, <https://arxiv.org/abs/1512.03385> (accessed Nov. 5, 2023).
- [61] “ResNet-50: The basics and a quick tutorial,” Datagen, <https://datagen.tech/guides/computer-vision/resnet-50/#:~:text=ResNet%2D50%20is%20a%2050,50%20Transfer%20Learning%20with%20Keras> (accessed Nov. 5, 2023).
- [62] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv.org, <https://arxiv.org/abs/2010.11929> (accessed Nov. 9, 2023).
- [63] C. Bhalerao, “Vision transformers [ViT]: A very basic introduction,” Medium, <https://medium.com/data-and-beyond/vision-transformers-vit-a-very-basic-introduction-6cd29a7e56f3> (accessed Nov. 9, 2023).
- [64] “Análisis ROC,” IBM, <https://www.ibm.com/docs/es/spss-statistics/beta?topic=features-roc-analysis> (accessed Nov. 1, 2023).
- [65] “Curvas PR Y ROC,” Medium, <https://medium.com/bluekiri/curvas-pr-y-roc-1489fbd9a527> (accessed Nov. 12, 2023).
- [66] “What is the difference between precision-recall curve vs Roc-AUC curve?,” Data Science Duniya, <https://ashutoshtripathi.com/2022/01/09/what-is-the-difference-between-precision-recall-curve-vs-roc-auc-curve/> (accessed Nov. 1, 2023).

- [67] “Aperio® AT2 - Digitalizador DE portaobjetos de Microscopio by Leica Biosystems: Medicaexpo,” El marketplace B2B del equipamiento médico, <https://www.medicaexpo.es/prod/leica-biosystems/product-95735-653510.html> (accessed Nov. 12, 2023).
- [68] The Endless Applications of digital pathology2 (2017) BioSpectrum India. Available at: <https://www.biospectrumindia.com/features/21/9897/the-endless-applications-of-digital-pathology.html> (Accessed: 30 October 2023)
- [69] Lu, M.Y. et al. (2021) ‘Data-efficient and weakly supervised computational pathology on whole-slide images’, Nature Biomedical Engineering, 5(6), pp. 555–570. doi:10.1038/s41551-020-00682-w.
- [70] Home (no date) faisal.ai. Available at: <https://faisal.ai/> (Accessed: 30 October 2023).
- [71] “TCGA-KICH,” GDC, https://portal.gdc.cancer.gov/projects?filters=%7B%22op%22%3A%22and%22%2C%22content%22%3A%5B%7B%22op%22%3A%22in%22%2C%22content%22%3A%7B%22field%22%3A%22projects.project_id%22%2C%22value%22%3A%5B%22TCGA-KICH%22%5D%7D%7D%5D%7D (accessed Nov. 1, 2023).
- [72] “TGCA-LUAD,” GDC, https://portal.gdc.cancer.gov/projects?filters=%7B%22op%22%3A%22and%22%2C%22content%22%3A%5B%7B%22op%22%3A%22in%22%2C%22content%22%3A%7B%22field%22%3A%22projects.project_id%22%2C%22value%22%3A%5B%22TCGA-LUAD%22%5D%7D%7D%5D%7D (accessed Nov. 1, 2023).
- [73] G. Litjens et al., “1399 H&E-stained sentinel lymph node sections of breast cancer patients: The CAMELYON dataset,” GigaScience, vol. 7, no. 6, 2018. doi:10.1093/gigascience/giy065
- [74] R. Hong, W. Liu, D. DeLair, N. Razavian, and D. Fenyö, “Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models,” Cell Reports Medicine, vol. 2, no. 9, p. 100400, 2021. doi:10.1016/j.xcrm.2021.100400

- [75] S. Fremond and S. Andani, “Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts,” *The Lancet*, 2022. doi:[https://doi.org/10.1016/S2589-7500\(22\)00210-2](https://doi.org/10.1016/S2589-7500(22)00210-2)
- [76] D. A. Levine, “Integrated genomic characterization of endometrial carcinoma,” *Nature*, vol. 497, no. 7447, pp. 67–73, 2013. doi:[10.1038/nature12113](https://doi.org/10.1038/nature12113)
- [77] Y. Dou et al., “Proteogenomic characterization of endometrial carcinoma,” *Cell*, vol. 180, no. 4, 2020. doi:[10.1016/j.cell.2020.01.026](https://doi.org/10.1016/j.cell.2020.01.026)
- [78] “Aperio LV1 RoW - Real-time Digital Pathology System,” Leica Biosystems, <https://www.leicabiosystems.com/digital-pathology/scan/aperio-lv1/> (accessed Nov. 4, 2023).
- [79] B. Jiang et al., “Biopsy versus resection for the management of low-grade gliomas,” *The Cochrane database of systematic reviews*, [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6478300/: :text=These %20are %20biopsy %2C %20the %20surgical,\(severity\) %20of %20the %20tumor.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6478300/: :text=These%20are%20biopsy%2C%20the%20surgical,(severity)%20of%20the%20tumor.) (accessed Nov. 4, 2023).
- [80] Jesús, “Qué es imagenet y por qué necesitas Conocerlo,” DataSmarts, <https://www.datasmarts.net/que-es-imagenet-y-por-que-necesitas-conocerlo/> (accessed Nov. 5, 2023).
- [81] M. Macenko, A method for normalizing histology slides for quantitative analysis, <https://www.cs.unc.edu/mn/sites/default/files/macenko2009.pdf> (accessed Nov. 13, 2023).
- [82] “Google/VIT-base-patch16-224-in21k · hugging face,” google/vit-base-patch16-224-in21k · Hugging Face, <https://huggingface.co/google/vit-base-patch16-224-in21k> (accessed Nov. 13, 2023).

- [83] C. C. Atabansi et al., “A survey of transformer applications for histopathological image analysis: New developments and Future Directions - Biomedical Engineering Online,” BioMed Central, <https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-023-01157-0> (accessed Nov. 7, 2023).
- [84] Y. Zhang et al., “Deep learning-based methods for classification of microsatellite instability in endometrial cancer from he-stained pathological images,” *Journal of Cancer Research and Clinical Oncology*, vol. 149, no. 11, pp. 8877–8888, 2023. doi:10.1007/s00432-023-04838-4
- [85] *Who Classification of Female Genital Tumours*. Lyon: International Agency for Research on Cancer, 2020.
- [86] V. Bushaev, “Adam-latest trends in deep learning optimization,” Medium, <https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c> (accessed Nov. 12, 2023).