TESIS DOCTORAL

UCAM
UNIVERSIDAD CATÓLICA
DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Programa de Doctorado en Ciencias de la Salud

TÉCNICA DE APRENDIZAJE PROFUNDO PARA LA AYUDA AL DIAGNÓSTICO EN CITOLOGÍAS CERVICOVAGINALES

Autor:
Ana Ortiz González

Directores:
Dr. D. Andrés Bueno Crespo
Dr. D. Juan Pedro Martínez Cendán

Murcia, Septiembre de 2025

TESIS DOCTORAL

UCAM
UNIVERSIDAD CATÓLICA
DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Programa de Doctorado en Ciencias de la Salud

TÉCNICA DE APRENDIZAJE PROFUNDO PARA LA AYUDA AL DIAGNÓSTICO EN CITOLOGÍAS CERVICOVAGINALES

Autor:
Ana Ortiz González

Directores:
Dr. D. Andrés Bueno Crespo
Dr. D. Juan Pedro Martínez Cendán

Murcia, Septiembre de 2025

AUTORIZACIÓN DEL DIRECTOR DE LA TESIS
PARA SU PRESENTACIÓN

El Dr. D. Andrés Bueno Crespo y el Dr. D. Juan Pedro Martínez Cendán como Directores[1] de la Tesis Doctoral titulada "TÉCNICA DE APRENDIZAJE PROFUNDO PARA LA AYUDA AL DIAGNÓSTICO EN CITOLOGÍAS CERVICOVAGINALES" realizada por D. Ana Ortiz González en el Programa de Doctorado en Ciencias de la Salud, **autoriza su presentación a trámite** dado que reúne las condiciones necesarias para su defensa.

Lo que firmo, para dar cumplimiento al Real Decreto 99/2011 de 28 de enero, en Murcia a 08 de septiembre de 2025.

ANDRES|BUENO|CRESPO

Firmado digitalmente por ANDRES|BUENO|CRESPO
Nombre de reconocimiento (DN): cn=ANDRES|BUENO|CRESPO, serialNumber=33382858Z, givenName=ANDRES, sn=BUENO CRESPO, ou=CIUDADANOS, o=ACCV, c=ES
Fecha: 2025.09.08 10:16:57 +02'00'

MARTINEZ CENDAN JUAN PEDRO - 22942847W

Firmado digitalmente por MARTINEZ CENDAN JUAN PEDRO - 22942847W
Fecha: 2025.09.16 18:36:26 +02'00'

RESUMEN

La citología cervicovaginal es uno de los métodos de elección en el diagnóstico precoz del cáncer de cérvix. El cáncer de cérvix representa la cuarta neoplasia más frecuente en mujeres a nivel mundial y afecta a todos los países del mundo. El empleo de la citología cervicovaginal como método de cribado busca la detección de lesiones preneoplásicas o displásicas, con el fin de dar un tratamiento precoz y frenar su progresión a lesiones de mayor agresividad biológica.

Debido a la creciente demanda de patólogos por el gran volumen de citologías, el examen manual puede ser tedioso y está sujeto a errores. El uso de sistemas de diagnóstico estandarizado como el *Sistema Bethesda de diagnóstico de la citología cervicovaginal* o la automatización del procesamiento en el laboratorio, son medidas que tratan de aligerar el proceso diagnóstico. El desarrollo de sistemas de detección automatizados del cáncer de cérvix por inteligencia artificial está en apogeo y está destacando como herramienta que permite optimizar los flujos de trabajo.

Planteamos el uso de modelos deep learning especializados en el diagnóstico por imagen, concretamente las redes neuronales convolucionales (CNN), aplicados al diagnóstico de células cervicales obtenidas de la citología cervicovaginal. Este trabajo desarrolla distintos métodos en busca de un modelo robusto que permita su implementación en la práctica diaria. Se desarrollan células artificiales como método de mejora de nuestra base de datos buscando una distribución de imágenes equilibrada entre las distintas categorías diagnósticas del sistema Bethesda. Examinamos si estas células artificiales son fieles a la célula original y son clasificadas por el modelo CNN en la categoría diagnóstica para la que fue diseñada. Exploramos distintos métodos que le aporten transparencia al modelo facilitando su comprensión y mejorando la confianza de los profesionales al empleo del mismo, como el uso de la segmentación con máscaras celulares en la entrada de la red como método de interpretabilidad (que justifica el funcionamiento o camino que toma la red para elaborar su predicción) o la generación automática de mapas

de calor en la capa de salida como método de explicabilidad (para justificar su decisión final). Por último, planteamos pequeñas modificaciones en el modelo CNN para su aplicación en países en vías de desarrollo con bajos recursos, como herramienta prometedora para el diagnóstico precoz del cáncer de cérvix, reduciendo las altas cifras morbilidad y mortalidad presente en estos países.

Los resultados en todas las vías de investigación son prometedores y le han aportado robustez al modelo. Sin embargo, se han puesto en evidencia una serie de limitaciones entre las que destaca la dificultad diagnóstica de las categorías intermedias del Sistema Bethesda; como ASC-US y LSIL, debido a la cierta ambigüedad que presentan en ocasiones sus características citológicas. Se plantean varias vías de solución y líneas futuras de investigación en busca de un modelo universal que se ponga en práctica en el trabajo asistencial.

PALABRAS CLAVE

Inteligencia artificial, citología, aprendizaje automático, cribado, cáncer de cérvix, Bethesda, displasia, CNN, explicabilidad, segmentación.

ABSTRACT

Cervicovaginal cytology is one of the methods of choice in the early diagnosis of cervical cancer. Cervical cancer is the fourth most frequent neoplasm in women worldwide and affects every country in the world. The use of cervicovaginal cytology as a screening method is aimed at detecting preneoplastic or dysplastic lesions, in order to provide early treatment and halt their progression to more biologically aggressive lesions.

Due to the increasing demand for pathologists because of the large volume of cytologies, manual examination can be tedious and subject to errors. The use of standardized diagnostic systems such as the Bethesda Diagnostic System for cervicovaginal cytology or the automation of laboratory processing are measures that attempt to streamline the diagnostic process. The development of automated cervical cancer detection systems using artificial intelligence is in full swing and is emerging as a tool to optimize workflows.

We propose the use of deep learning models specialized in diagnostic imaging, specifically convolutional neural networks (CNN), applied to the diagnosis of cervical cells obtained from cervicovaginal cytology. This work develops different methods in search of a robust model that can be implemented in daily practice. Artificial cells are developed as a method of improving our database, seeking a balanced distribution of images among the different diagnostic categories of the Bethesda system. We examine whether these artificial cells are faithful to the original cell and are classified by the CNN model in the diagnostic category for which it was designed. We explored different methods that bring transparency to the model, facilitating its understanding and improving the confidence of professionals in its use, such as the use of segmentation with cell masks at the network input as a method of interpretability (which justifies the operation or path taken by the network to make its prediction) or the automatic

generation of heat maps in the output layer as a method of explainability (to justify its final decision). Finally, we propose minor modifications to the CNN model for its application in low-resource developing countries as a promising tool for the early diagnosis of cervical cancer, reducing the high morbidity and mortality rates present in these countries.

The results in all avenues of research are promising and have provided robustness to the model. However, a series of limitations have been highlighted, including the diagnostic difficulty of the intermediate categories of the Bethesda System, such as ASC-US and LSIL, due to the certain ambiguity that their cytological characteristics sometimes present. Several solutions and future lines of research are proposed in search of a universal model to be put into practice in health care work.

KEYWORDS

Artificial intelligence, cytology, machine learning, screening, cervical cancer, Bethesda, dysplasia, CNN, explainability, segmentation.

AGRADECIMIENTOS

Gracias a los que me han animado a realizar este trabajo, a los que me han ayudado a llevarlo a cabo y a los que me han enseñado que la inteligencia artificial es sorprendente.

ÍNDICE GENERAL

RESUMEN **7**

I - INTRODUCCIÓN **19**

1.1. EL CÉRVIX UTERINO 19
1.1.1. 1.1. ANATOMÍA DEL CÉRVIX UTERINO 19
1.1.2. HISTOLOGÍA DEL CÉRVIX UTERINO 22
1.1.3. EXPLORACIÓN CLÍNICA DEL ORIFICIO CERVICAL EXTERNO 27
1.2. EL CÁNCER DE CÉRVIX 28
1.2.1. SITUACIÓN DEL CÁNCER DE CÉRVIX A NIVEL MUNDIAL 28
1.2.2. SITUACIÓN DEL CÁNCER DE CÉRVIX EN ESPAÑA 33
1.2.3. ANTECEDENTES Y ESTUDIO COLPOSCÓPICO DEL CÁNCER DE CÉRVIX 35
1.2.4. NOMENCLATURA DE LA DISPLASIA CERVICAL: TERMINOLOGÍA LAST 36
1.2.5. HISTOPATOLOGÍA DE LAS LESIONES PRECURSORAS DEL CÁNCER DE CÉRVIX 38
1.2.6. ESTRATEGIA DE CRIBADO DEL CÁNCER DE CÉRVIX 44
1.3. VIRUS DEL PAPILOMA HUMANO Y SU INFLUENCIA SOBRE EL CÁNCER DE CÉRVIX 47
1.3.1. GENOTIPOS DEL VIRUS DEL PAPILOMA HUMANO 47
1.3.2. EPIDEMIOLOGÍA 48
1.3.3. ETIOPATOGENIA 49
1.4. LA CITOLOGÍA CERVICOVAGINAL 51
1.4.1. HISTORIA Y FUNDAMENTOS BÁSICOS DE LA CITOLOGÍA CERVICOVAGINAL 51
1.4.2. ELABORACIÓN DE LA CITOLOGÍA CERVICOVAGINAL 54
1.4.3. DIAGNÓSTICO DE LA CITOLOGÍA CERVICOVAGINAL CON EL SISTEMA BETHESDA 58
1.4.4. LIMITACIONES Y FRONTERAS EN EL DIAGNÓSTICO CITOLÓGICO 64
1.5. ESTADO DEL ARTE DE LA INTELIGENCIA ARTIFICIAL 65
1.5.1. HISTORIA Y CRONOLOGÍA EN LOS SISTEMAS DE LECTURA AUTOMATIZADA DE CITOLOGÍAS CERVICOVAGINALES 65
1.5.2. COMIENZOS DE LA INTELIGENCIA ARTIFICIAL 68
1.5.3. FUNCIONAMIENTO DEL DEEP LEARNING 72
1.6. LA APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL EN ANATOMÍA PATOLÓGICA 75
1.6.1. DIGITALIZACIÓN E INICIOS DE LA PATOLOGÍA DIGITAL 75
1.6.2. LOS ESCÁNERES EN PATOLOGÍA DIGITAL 77
1.6.3. HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL DE AYUDA AL DIAGNÓSTICO EN ANATOMÍA PATOLÓGICA 78
1.6.4. TÉCNICAS DE MACHINE LEARNING EN PATOLOGÍA 81
1.6.5. REDES NEURONALES CONVOLUCIONALES APLICADAS AL DIAGNÓSTICO DE PREPARACIONES DIGITALES 82
1.6.6. RETOS Y BARRERAS EN LA APLICACIÓN DE LA IA EN ANATOMIA PATOLÓGICA 85
1.7. LA IA EN CITOPATOLOGÍA 87

1.7.1. FUNDAMENTO Y BASES DEL DESARROLLO DE LA IA EN CITODIAGNÓSTICO 87

1.7.2. ANTECEDENTES DEL EMPLEO DE CNN COMO MODELO DE CLASIFICACIÓN CELULAR EN CITOLOGÍAS 89
1.7.3. SISTEMAS DE DETECCIÓN DISPONIBLES EN LA ACTUALIDAD 99
1.7.4. FASES EN LA IMPLANTACIÓN DE UN DIAGNÓSTICO AUTOMATIZADO 104
1.7.5. RETOS DEL CITODIAGNÓSTICO AUTOMATIZADO 105

II - JUSTIFICACIÓN 109

III - OBJETIVOS 117

3.1. OBJETIVO PRINCIPAL 117
3.2. OBJETIVOS SECUNDARIOS 117

IV - MATERIAL Y MÉTODO 121

4.1. BASE DE DATOS 121
4.1.1. BASE DE DATOS AL INICIO DEL PROYECTO 121
4.1.2. BASE DE DATOS MEJORADA CON CÉLULAS ARTIFICIALES 126
4.2. CREACIÓN DE CÉLULAS ARTIFICIALES 127
4.3. ELABORACIÓN DE UN MODELO HÍBRIDO PARA LA CLASIFICACIÓN DE IMÁGENES 130
..... 132
4.4. LA SEGMENTACIÓN COMO AYUDA AL DIAGNÓSTICO POR IMAGEN. 133
4.4.1. PREPARACIÓN DE NUESTRA BASE DE DATOS PARA SEGMENTACIÓN 135
4.4.2. DISEÑO DE NUESTRO MODELO DE SEGMENTACIÓN Y CLASIFICACIÓN CELULAR 136
4.4.3. EL EMPLEO DE DISTINTOS CLASIFICADORES EN LOS MODELOS DE SEGMENTACIÓN Y CLASIFICACIÓN CELULAR PROPUESTOS 140
4.5. LA EXPLICABILIDAD EN LOS MODELOS DE DIAGNÓSTICO POR IMAGEN 144
4.5.1. APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL EXPLICABLE EN CITOLOGÍAS CERVICOVAGINALES 146
4.5.2. DISEÑO DE UN MODELO HÍBRIDO EXPLICABLE MEDIANTE EL USO DE UNA RED GRAD-CAM CON MAPAS DE CALOR 150
4.5.3. BASE DE DATOS EMPLEADA Y SIMPLIFICACIÓN DE LOS RESULTADOS PARA ANALIZAR LA UTILIDAD DEL MODELO 152
4.6. EL USO DE LA INTELIGENCIA ARTIFICIAL COMO AYUDA AL DIAGNÓSTICO EN PAÍSES EN VÍAS DE DESARROLLO 154
4.6.1. PROPUESTA DE UN MODELO DE BAJO COSTE EN PAÍSES EN VÍAS DE DESARROLLO 155

V - RESULTADOS 159

5.1. NUEVA BASE DE DATOS MEDIANTE LA CREACIÓN DE CÉLULAS ARTIFICIALES 159
5.2. NUESTRO MODELO HÍBRIDO DCGAN EN LA CLASIFICACIÓN CÉLULAS ESCAMOSAS CERVICALES 160
5.3. LA SEGMENTACIÓN COMO MEJORA DEL MODELO DE DIAGNÓSTICO CELULAR 167
5.3.1. ELABORACIÓN AUTOMÁTICA DE MÁSCARAS DE SEGMENTACIÓN 167

5.3.2. CLASIFICACIÓN CELULAR TRAS LA SEGMENTACIÓN 168
5.4. EL IMPACTO DE LA EXPLICABILIDAD EN LA CLASIFICACIÓN CELULAR 179
5.4.1. Primera etapa de explicabilidad 179
5.4.2. Segunda etapa de clasificación. 181
5.5. EL USO DE LA IA COMO HERRAMIENTA DIAGNÓSTICA EN PAÍSES EN VÍAS DE DESARROLLO 186
5.5.1. MANEJO DE LA IMAGEN OBTENIDA POR UNA VIDEO-CAPTURADORAS DE MICROSCOPIO 187
5.5.2. MANEJO DE LA IMAGEN OBTENIDA POR DISPOSITIVOS DE BAJO COSTE 189
5.5.3. DISEÑO DE UN MODELO DE BAJO COSTE 191

VI - DISCUSIÓN 199

6.1. EL USO DE IMÁGENES ORIGINALES NO PREPROCESADAS 199
6.2. MEJORA DE LA NUEVA BASE DE DATOS MEDIANTE LA CREACIÓN DE CÉLULAS ARTIFICIALES 200
6.3. POTENCIAL DE NUESTRO MODELO HÍBRIDO DCGAN EN EL DIAGNÓSTICO DE CÉLULAS CERVICALES 201
6.4. SEGMENTACIÓN COMO MÉTODO DE INTERPRETABILIDAD 207
6.5. LOS MAPAS DE CALOR COMO MÉTODO DE EXPLICABILIDAD 218
6.6. MEJORA DE LA CALIDAD DE LA IMAGEN DE DISPOSITIVOS DE BAJO RENDIMIENTO Y DISEÑO DE MODELOS DE BAJO COSTE EN PAÍSES EN VÍAS DE DESARROLLO 222

VII - CONCLUSIONES 227

VIII - LIMITACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN 231

8.1. LIMITACIONES 231
8.2. FUTURAS LÍNEAS DE INVESTIGACIÓN 233
8.3. CRITERIOS DE CUANTIFICACIÓN A LA RED PARA TRATAR DE DISCERNIR ENTRE CATEGORÍAS INTERMEDIAS 234
..... 235
8.4. CRITERIOS CUALITATIVOS QUE ANALIZA LA RED CON EL ETIQUETADO CELULAR MANUAL COMO MÉTODO DE RAZONAMIENTO DIAGNÓSTICO. 236

IX - REFERENCIAS BIBLIOGRÁFICAS 241

X - ANEXOS 257

SIGLAS Y ABREVIATURAS

AEPCC. Asociación Española de Patología Cervical y Colposcopia.

ASC-US. Atipia de células escamosas de significado incierto.

CAD. Diagnóstico Médico asistidos por Computador o ***Computer Assited Diagnóstic systems.***

CCV. Citología cervicovaginal.

CNN. Redes neuronales convolucionales.

DCGAN. D*eep Convolutional Geberative Adversarial Network*.

DCNN. Redes neuronales convolucionales profundas.

DL. ***Deep Learning*.**

FCN. Red totalmente convolucional.

FDA. ***Food and Drug Administration*.**

GAN. Redes generativas adversarias.

GAP. ***Global Average Pooling o*** **agrupación media global.**

BG. Gigabytes.

GLOBOCAN. ***Global Cancer Observatory*.**

Grad-CAM. Mapeo por activación de clases ponderado por gradiente

HPV. ***Human Papilloma Virus*.**

HSIL. ***High grade squamous intraepitelial lesion.***

IA. Inteligencia Artificial.

IARC. ***International Agency for Research on Cancer.***

LAST. ***Low Anogenital Squamous Terminology*.**

LSIL. ***Low grade squamous intraepitelial lesion.***

ML. ***Machine Learning*.**

NCI. ***National Cancer Institute.***

OCE. Orificio cervical externo.

OCI. Orificio cervical interno.

OMS. Organización Mundial de la Salud.

PCR. Reacción en cadena de la polimerasa.

PD. Patología Digital.

pRT. Retinoblastoma.

Real-ESRGAN. E*nhanced super-resolution generative adversarial network.*

ResNet. *Residual Network*

SEOM. Sociedad Española de Oncología Médica.

SIL. *Squamous intraepitelial lesion.*

UEC. Unión escamocolumnar.

VIH. Virus de la inmunodeficiencia humana.

WSI. W*hole slide image.*

XIA. Inteligencia artificial explicable.

ZT. Zona de transición.

ÍNDICE DE FIGURAS, DE TABLAS Y DE ANEXOS

ÍNDICE DE FIGURAS

Figura 1.1. En esta imagen podemos observar las relaciones anatómicas del cérvix con el cuerpo del útero, anejos y resto de vísceras pélvicas. Imagen extraída del libro Atlas de Anatomía humana, de Frank H. Netter (6ª Edición). 19

Figura 1.2. Espéculo que permite realizar una exploración ginecológica completa del cuello del útero mediante la dilatación mecánica de las paredes de la vagina, haciendo al orificio cervical externo completamente visible. Fotografía cortesía de la Dra. Ortiz González del Servicio de Oncología Radioterápica del Hospital Universitari Son Espases (Palma)..... 20

Figura 1.3. Imagen de las distintas partes del cérvix. Extraída del NCI (National Cancer Institute). 21

Figura 1.4. Exploración ginecológica del cuello del útero. Podemos observar al fondo la mucosa cervical y el orificio cervical externo, que da paso al canal que comunica el cérvix con el útero, conocido como canal endocervical. Fotografía cortesía de la Dra. Lara Peñaranda del Servicio de Ginecología y Obstetricia del Hospital Santa Lucía (Cartagena). 21

Figura 1.5. Imagen microscópica de la mucosa exocervical. Nótese los distintos estratos del epitelio escamoso que la conforman: el estrato basal, intermedio y superficial. Elaboración propia. 24

Figura 1.6. Imagen microscópica de la unión escamocolumnar en la que se observa la transición entre ambos epitelios. A la izquierda se evidencia el epitelio escamoso exocervical y a la derecha el epitelio cilíndrico simple endocervical. Elaboración propia. 26

Figura 1.7. Imagen microscópica de la mucosa endocervical revestida por un epitelio cilíndrico simple mucosecretor. Elaboración propia. 26

Figura 1.8. Imagen esquemática de cómo va evolucionando la unión escamocolumnar en la exploración clínica del cérvix a lo largo de los años. Imagen obtenida de "La colposcopia y el tratamiento de la neoplasia intraepitelial cervical: Manual para principiantes" de Sellors y Sankaranarayanan, modificada por Dr. Remezal Solano del Hospital Universitario Virgen de la Arrixaca (Murcia). 27

Figura 1.9. Casos de cáncer a nivel mundial en ambos sexos registrados en el año 2020. Gráfico obtenido de GLOBOCAN. 30

Figura 1.10. Casos de cáncer a nivel mundial en mujeres registrados en el año 2022. Gráfico obtenido de GLOBOCAN. 30

Figura 1.11. Datos de mortalidad por cáncer en mujeres en el 2022, donde se observan una mayor distribución de los fallecimientos por cáncer de cérvix en países de África y América del Sur. Mapamundi obtenido de GLOBOCAN..... 31

Figura 1.12. Números absolutos de mortalidad atribuidas al cáncer en mujeres durante el año 2022. El cáncer de cérvix ocupa la cuarta posición con un 8,1% de los fallecimientos. Gráfico obtenido de GLOBOCAN. 31

Figura 1.13. Incidencia en España de distintos tipos de cáncer por cada 100.000 mujeres / año, según los datos recogidos en GLOBOCAN en 2022. Gráfico obtenido de GLOBOCAN. 33

Figura 1.14. Incidencia por cada 100.000 mujeres del cáncer de cérvix recogidos en el año 2022 en todo el mundo. Mapamundi obtenido de GLOBOCAN. 34

Figura 1.15. Colposcopia con test de Schiller positivo, con el que se evidencia un área color mostaza pobre en glucógeno sospechosa de displasia de la mucosa cervical normal rica en glucógeno marcada de marrón. Fotografía cortesía de la Dra. Lara Peñaranda del Servicio de Ginecología y Obstetricia del Hospital Santa Lucía (Cartagena). 35

Figura 1.16. Imagen microscópica de una lesión intraepitelial escamosa de bajo grado (LSIL). Nótese la presencia de alteraciones citoarquitecturales en capas bajas del epitelio (arriba) y la presencia de coilocitos en capas altas del epitelio (abajo). Elaboración propia. 41

Figura 1.17. Imagen microscópica de una lesion intraepitelial escamosa de alto grado HSIL. Observamos que las células presentan alteraciones citoarquitecturales que ocupan estratos intermedios del epitelio compatibles con un CIN 2 (arriba) y superficiales compatibles con CIN 3 (abajo). Elaboración propia. 42

Figura 1.18. Detalle del HSIL por CIN 3. Nótese la intensa alteración citoarquitectural que ocupa por completo el epitelio con alteración de la polaridad nuclear, aumento de la relación núcleo-citoplasma, presencia de pleomorfismo nuclear y mitosis. Elaboración propia. 43

Figura 1.19. Doble tinción inmunohistoquímica con p16/ki67. Esta tinción marca de color marrón el p16 que señala el citoplasma de las células infectadas por el HPV y de color rojo el núcleo de células proliferativas, con que muestra positividad más allá de la capa parabasal. Podemos interpretar que estas células están infectadas y que presentan cambios displásicos que llegan a capas altas del epitelio, compatible con un HSIL por CIN 3. Elaboración propia. 43

Figura 1.20. Ciclo infectivo del HPV. Imagen extraída de la comunicación "Aplicación de técnicas complementarias a la citología, para la detección del virus del papiloma humano (HPV) en lesiones epiteliales del aparato genital femenino", de García Macías.y cols. del V Congreso Virtual Hispanoamericano de Anatomía Patológica..... 50

Figura 1.21. Comparativa de distintos métodos de elaboración de la citología. (A) Pipeta y botella que se usaban para la elaboración de la citología cervicovaginal descrita por Papanicolau, extraída de la publicación "The diagnostic value of vaginal smears in carcinoma of the uterus" (Papanicolaou & Traut, 1997). (B) Método actual de toma de citología líquida que emplea un cepillo cervical y líquido conservante celular, extraída de la web "Artemedica" (disponible en: https://artemedica.es/ginecologia/citologia-liquida-un-mejor-diagnostico-en-cancer-de-cuello-de-utero/). 52

Figura 1.22. Método de recogida de muestra mediante la triple toma o citología convencional. De forma separada se toma la muestra del endocérvix (e), exocérvix (c) y vagina (v) y se depositan de manera independiente en 3 territorios del portaobjetos bien delimitados. Imagen obtenida de https://citomol.com/citologias/. 54

Figura 1.23. Comparativa del portaobjetos con muestra obtenida mediante citología líquida (arriba) y triple toma (abajo). Nótese que en la citología líquida todas las células de todos los territorios muestreados se disponen juntos en un área redondeada bien delimitada. Por el contrario, en la triple toma, se observan distintas secciones lineales en el cristal, cada una correspondiente a cada uno de los territorios muestreados de manera independiente. Elaboración propia. 56

Figura 1.24. Líquido conservante celular que se emplea para la fijación de la muestra en la citología liquida junto a su cepillo especial para la recogida de la muestra. Elaboración propia. 56

Figura 1.25. Maquinara SuperPap® empleada en la elaboración automatizada de la citología líquida. Elaboración propia. 57

Figura 1.26. Células escamosa cervicales benignas. Nótese una baja relación de la relación núcleo citoplasma. Podemos distinguir células exocervicales de estratos superficiales (de coloración orangófila o rosada) e intermedios (de coloración cianófila o azulada). Elaboración propia. 59

Figura 1.27. Células con ligero aumento de la relación núcleo-citoplasma e irregularidad nuclear ligera, que no son suficientes para catalogarlo de una lesión de mayor agresividad. Elaboración propia. 60

Figura 1.28. Placa de coilocitos LSIL en la que se observa binucleación, aumento de la relación núcleo-citoplasma con irregularidad e hipercromasia nuclear, con un halo blanquecino perinuclear amplio, que define a los coilocitos. Elaboración propia. 62

Figura 1.29. Se observa una célula epitelial de tipo intermedia con alteraciones citológicas marcadas con un aumento de la relación núcleo-citoplasma e irregularidad nuclear. Elaboración propia. 62

Figura 1. 30. Se observa en el centro dos células epiteliales del estrato basal o parabasal con un aumento marcado del tamaño nuclear, con intenso hipercromatismo e irregularidad nuclear. Elaboración propia. 63

Figura 1.31. Placa de células con HSIL. Elaboración propia. 63

Figura 1.32. Flujo de trabajo que muestra la inserción de AutoPap 300 QC en el flujo de trabajo del laboratorio del hospital. Elaboración propia. 67

Figura 1.33. Monitor de visionado de las células potencialmente anormales del sistema PapNet. Imagen obtenida de la revisión sistemática: "A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis" de Jiang y cols.(Jiang et al., 2023). 67

Figura 1.34. Relación entre la inteligencia artificial, el aprendizaje automático y el aprendizaje profundo. Imagen tomada de Marcial García Rojo Aplicaciones de deep learning, un método de inteligencia artificial, en Anatomía Patológica (Libro Blanco 2021). 70

Figura 4.1. Flujo de trabajo para la elaboración y escaneado de una citología cervicovaginal. A.- Sistema de recogida de la muestra del paciente mediante un cepillado cervical que se remite a un fijador celular BD SurePath® liquid Pap test®. B.- Procesado automático citológico mediante el sistema SuperPap®. C.- Escáner Ventana DP200 de escaneado de laminillas. D.- Imagen digital obtenida de todo el portaobjetos (whole slide image, WSI) sobre la que podemos navegar y modificar su aumento para el estudio celular a alta definición. Elaboración propia. 122

Figura 4.2. Imagen digital de una citología cervicovaginal escaneada con el escáner Ventana DP200 a un aumento digital 20x. Elaboración propia. 123

Figura 4.3. Imágenes individuales recortadas del escaneado de todo el portaobjetos. Estas células fueron clasificadas por un patólogo experto en las siguientes categorías (de izquierda a derecha): célula benigna, ASC-US, LSIL y HSIL. Imagen tomada de la tesis doctoral "Clasificación automática de tumores de ovario y citologías cervicovaginales a partir de imágenes ecográficas y microscópicas mediante su análisis con técnicas de aprendizaje automático" del Dr. Martínez Más. 124

Figura 4.4. Ejemplos en los que la imagen de la célula es de difícil valoración por presentar solapamientos, plegamientos o borrosidad. De izquierda a derecha: célula benigna, ASC-US, LSIL y HSIL. Imagen tomada de la tesis doctoral "Clasificación automática de tumores

de ovario y citologías cervicovaginales a partir de imágenes ecográficas y microscópicas mediante su análisis con técnicas de aprendizaje automático" del Dr. Martínez Más. 125

Figura 4.5. Conjunto de células artificiales generadas por la red StyleGAN2 que incluyen tanto las células correctamente generadas como las mal generadas: (A) benignas, (B) LSIL, (C) HSIL y (D) ASC-US. Elaboración propia. 129

Figura 4.6. Modelo híbrido propuesto en el que se combina una red GAN y una DCNN para obtener una nueva base de datos amplia y mejorada. Elaboración propia. 132

Figura 4.7. Estructura del modelo híbrido DCGAN. Elaboración propia..... 133

Figura 4.8. Fases 1 y 2 del diseño del modelo de segmentación automática. En primer lugar se prepara una base de datos de células segmentadas manualmente para entrenar a la red U-Net. Posteriormente se utilizan nuevas imágenes no testeadas previamente como imagen de entrada a la espera de obtener una célula con una máscara generada automáticamente en la capa de salida. Elaboración propia. 138

Figura 4.9. Fase 3 del modelo propuesto. Una vez que la red aprende a segmentar automáticamente, esta se fusiona con una red Xception mediante un clasificador. Dependiendo del modelo de clasificador, proponemos distintos modelos. Elaboración propia. 140

Figura 4.10. La capa aplanada convierte la información en un vector plano que presenta las características necesarias para su lectura por la capa densa que tiene todas las redes conectadas entre sí. Elaboración propia. 142

Figura 4.11. Capa GAP como clasificador. Realiza un valor promedio de los mapas de características generados reduciendo significativamente la cantidad de datos. Elaboración propia. 144

Figura 4.12. Arquitectura propuesta por el modelo compuesto de 2 redes CNN tipo Xception. El modelo de la izquierda permite extraer el mapa de calor que se superpone a la imagen original en función de un porcentaje de ponderación marcado por el parámetro α. Esta imagen fusionada sirve de insumo al modelo de la derecha que clasifica a la célula en una categoría diagnóstica. Elaboración propia. 151

Figura 5.1. Imágenes generadas por la GAN para cada categoría diagnóstica: (a) HSIL, (b) ASC-US, (c) LSIL y (d) benigna. Elaboración propia..... 161

Figura 5.2. Porcentaje de acierto en la elaboración y etiquetado de células artificiales del modelo DCGAN de cada categoría diagnóstica. Elaboración propia. 161

Figura 5.3. Imágenes que comparan la segmentación manual con la imagen real. A la izquierda una célula benigna y a la derecha una célula H-SIL con sus respectivas segmentaciones, marcando en rojo los núcleos y en azul los citoplasmas. Elaboración propia. 167

Figura 5.4. Resultado de la segmentación automática mediante algoritmos de deep learning. A la izquierda una célula benigna y a la derecha HSIL, diferenciando el núcleo en rojo, el citoplasma en verde y el fondo el azul. Elaboración propia. 168

Figura 5.5. Modelos propuestos, donde la U-Net aprende a segmentar y se fusiona con una red Xception mediante un clasificador. Marcado en verde el contorno de la arquitectura de los modelos 1 y 2 y en rojo del modelo clásico. Elaboración propia (Ortiz-González et al., 2025)..... 169

Figura 5.6. Resultados de la clasificación del modelo en 4 clases, comparando el modelo propuesto 1, 2 y el modelo clásico de referencia. Del inglés, "AUC" hace referencia al área bajo la curva, "F1-score" al valor F1, "recall" a la sensibilidad o tasa de verdaderos

positivos, "precision" a precisión y "accuracy" a exactitud. Elaboración propia (Ortiz-González et al., 2025). 171

Figura 5.7. Resultados de la clasificación del modelo en 2 clases en el que se comparan las métricas para los 3 modelos a estudio. Del inglés, "AUC" hace referencia al área bajo la curva, "F1-score" al valor F1, "recall" a la sensibilidad o tasa de verdaderos positivos, "precision" a precisión y "accuracy" a exactitud. Elaboración propia (Ortiz-González et al., 2025)..... 173

Figura 5.8. Matriz de confusión del mejor modelo propuesto (modelo 2) para analizar el rendimiento del modelo en una clasificación binaria (a) y multiclase en 4 categorías diferentes (b). Elaboración propia (Ortiz-González et al., 2025). 175

Figura 5.9. Resultados tras la fusión de las imágenes reales con los mapas de calor para distintos valores del parámetro α. Se resaltan los valores de α = 0.1 y 0.3 para ver con más detalle la influencia del mapa de calor. Elaboración propia. 180

Figura 5.10. Métricas de los resultados obtenidos para la exactitud (arriba a la izquierda), precisión (arriba a la derecha), exhaustividad (abajo a la izquierda) y parámetro F1 (abajo a la derecha). En el eje Y se observa el valor de la métrica a estudio y en el eje X el valor del parámetro α. Cada línea de color determina el valor del parámetro a estudio para un determinado valor α, que presentan un punto central que señala la media de 10 determinaciones realizadas para cada valor de α. Elaboración propia..... 183

Figura 5.11. Comparativa de un campo digital de células obtenidas mediante una cámara low cost (arriba) frente a una laminilla escaneada con el escáner Ventana DP200 (abajo). Nótese la aceptable calidad de la imagen low-cost que; sin embargo, resulta mejorable al compararla con la imagen digital. Elaboración propia. 187

Figura 5.12. A la izquierda vemos un microscopio trinocular con una cámara video-capturadora low-cost instalada en su parte superior. A la derecha vemos el software gratuito que nos permite ver la imagen en tiempo real. Elaboración propia. 188

Figura 5.13. Algoritmo de mejora de la resolución de la imagen. A la izquierda una imagen del dispositivo low-cost original. A la derecha la misma imagen mejorada. Elaboración propia. 188

Figura 5.14. Imágenes de baja y media calidad de una célula ASC-US sin mejorar (arriba) y mejorada (abajo). Imagen obtenida del trabajo fin de grado "Sistema de diagnóstico para el cáncer de cérvix mediante dispositivos de bajo coste" de D. Juan Miguel Correa García. Elaboración propia. 190

Figura 5.15. Imágenes de distintas calidades mezcladas sin mejorar (arriba) y mejoradas (debajo). Imagen obtenida del trabajo fin de grado "Sistema de diagnóstico para el cáncer de cérvix mediante dispositivos de bajo coste" de D. Juan Miguel Correa García. Elaboración propia. 190

Figura 5.16. Modelo low-cost propuesto para el diagnóstico de citología cervicovaginal. Elaboración propia. 192

Figura 5.17. Resultados de comparar el modelo original (barra en azul), modelo TensorFlow lite (barra en rojo) y TensorFlow lite cuantizado (barra en verde). Elaboración propia. .. 194

Figura 5.18. Modelo de explicabilidad mediante la fusión de la célula original y un mapa de calor. Elaboración propia. 195

Figura 6.1. Imágenes de células artificiales cuya superposición celular, falta de nitidez o exceso de tinción podría justificar un etiquetado erróneo a HSIL, siendo originalmente creadas como LSIL (fila de arriba) y ASC-US (fila de abajo). Elaboración propia..... 204

ÍNDICE DE TABLAS

Tabla 1.1. Tabla comparativa con la evolución de la nomenclatura de la displasia cervical con los años y cuáles son las equivalencias entre las distintas clasificaciones. Elaboración propia. 37

Tabla 4.1. Base de datos empleada para elaborar el modelo explicable. Elaboración propia. 153

Tabla 5.1. Distribución de las imágenes de las células reales obtenidas y las células artificiales generadas, entre las distintas categorías diagnósticas del Sistema Bethesda. Elaboración propia. 159

Tabla 5.2. Matriz de confusión en la que se comparan los resultados reales generados por un patólogo experto con los obtenidos de forma automática por el modelo DCGAN. Las filas determinan los diagnósticos para los que la célula artificial fue generada (2000 por cada categoría diagnóstica) y en las columnas la etiqueta generada por el modelo. Elaboración propia. 160

Tabla 5.3. Tasa de verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN) y falsos negativos (FN) obtenidos por el modelo de clasificación de células artificiales. Elaboración propia. 162

Tabla 5.4. Análisis de métricas de sensibilidad (S), especificidad (E), precisión (P), valor predictivo positivo (VPN), exactitud (del inglés accurancy, ACC), valor F1 (F1), tasa de falsos positivos (TFP) y tasa de falsos negativos (TFN). Elaboración propia. 164

Tabla 5.5. Tasa de predicción del modelo DCGAN de las 2000 células artificiales generadas para cada categoría diagnóstica. Se detalla el porcentaje de células artificiales clasificada en cada categoría diagnóstica por el modelo. Elaboración propia. 165

Tabla 5.6. Matriz de confusión tras la reclasificación de los resultados en "revisables" (ASC-US, LSIL y HSIL) y "no revisables"(benignas) comparando las predicciones de la red frente a la clasificación real. En las filas se detalla la clasificación real y en las columnas la predicción automática del modelo. Elaboración propia. 166

Tabla 5.7. Porcentaje de células etiquetadas como "revisables" y "no revisables" tras su reclasificación binaria del total de 2000 células "no revisables" y 6000 "revisables". Elaboración propia. 166

Tabla 5.8. Tabla comparativa de las métricas de los 3 modelos estudiados, enfrentando la clasificación multiclase con la binaria. Elaboración propia..... 173

Tabla 5.9. Tabla de las métricas obtenidas tras el análisis de los resultados de la matriz de confusión de la clasificación multiclase del modelo 2. La muestra consta de 30 células en cada categoría, con un total de 120. Se estudian: verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN), verdaderos negativos (VN), sensibilidad, especificidad, precisión , valor predictivo negativo, exactitud, valor F1, tasa de falsos positivos y tasa de falsos negativos. Elaboración propia. 176

Tabla 5.10. Tabla con las métricas obtenidas tras el análisis de la matriz de confusión de la clasificación celular binaria del modelo 2. La muestra consta de 90 células revisables y 30 no revisables; alcanzando una muestra total de 120 células. Se estudian: verdaderos

positivos (VP), falsos positivos (FP), falsos negativos (FN), verdaderos negativos (VN), sensibilidad, especificidad, precisión , valor predictivo negativo, exactitud, valor F1, tasa de falsos positivos y tasa de falsos negativos. Elaboración propia..... 177

Tabla 5.11. Distribución de la clasificación celular comparando el sistema binario y el multiclase del modelo 2. Elaboración propia..... 178

Tabla 5.12. Porcentaje de células clasificadas correctamente en cada categoría de la predicción multiclase del modelo 2. Nótese en las celdillas marcadas en verde las células correctamente clasificadas, que representan la sensibilidad de la red para cada categoría. Elaboración propia. 178

Tabla 5.13. Matriz de confusión que compara los resultados obtenidos, empleando imágenes con $\alpha = 0.0$ y $\alpha = 0.1$. Se detalla la exactitud o precisión global para cada valor del parámetro α. Elaboración propia..... 184

Tabla 5.14. Tabla con los parámetros obtenidos a través de la matriz de confusión de los resultados obtenidos para $\alpha = 0.0$ y $\alpha = 0.1$. Elaboración propia..... 185

Tabla 5.15. Se detalla el porcentaje de células bien etiquetadas y mal etiquetadas en una clasificación binaria ("revisables" y "no revisables"), después del uso de imágenes originales ($\alpha = 0.0$) y las imágenes fusionadas con α =0.1. Elaboración propia. 186

Tabla 8.1. Estimación de la frecuencia de cada diagnóstico en la citología cervicovaginal. Son medidas estimadas obtenidas de estudios epidemiológicos y guías de cribado internacionales. Elaboración propia. 232

Tabla 8.2. Matriz de confusión para la clasificación multiclase de categorías individuales (H-SIL, L-SIL, ASC-US y Normal) utilizando las imágenes originales como entrada, resaltando en negrita las clasificaciones correctas en cada categoría. Tabla obtenida de la tesis doctoral del Dr. Martínez Más "Clasificación automática de tumores de ovario y citologías cervicovaginales a partir de imágenes ecográficas y microscópicas mediante su análisis con técnicas de aprendizaje automático"..... 235

ÍNDICE DE ANEXOS

Anexo 1. Enfoques de entrenamiento de CNN. Imagen obtenida de la revisión sistemática: "A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis" de Jiang y cols. (Jiang et al., 2023)..... 258

Anexo 2. Fases de implantación del diagnóstico automatizado en citología cervicovaginal. Imagen obtenida de la revisión sistemática: "A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis" de Jiang y cols. (Jiang et al., 2023)..... 261

I – INTRODUCCIÓN

I - INTRODUCCIÓN

1.1. EL CÉRVIX UTERINO

1.1.1. 1.1. ANATOMÍA DEL CÉRVIX UTERINO

El cérvix uterino o cuello del útero, constituye una estructura cilíndrica fibromuscular que conecta la cavidad uterina con la vagina. En su cara anterior se relaciona con la vejiga urinaria a través de un tejido conjuntivo de sostén laxo y en su cara posterior se recubre por peritoneo y se separa del recto por un pequeño espacio llamado fondo de saco rectouterino (Rouvière, 2005) (**Figura 1.1**). En condiciones normales, el cérvix tiene unas dimensiones aproximadas de 3-4 cm de longitud máxima y 3 cm de diámetro, que dependen de la edad de la paciente y de si ha sido nulípara o ha tenido varias gestaciones.

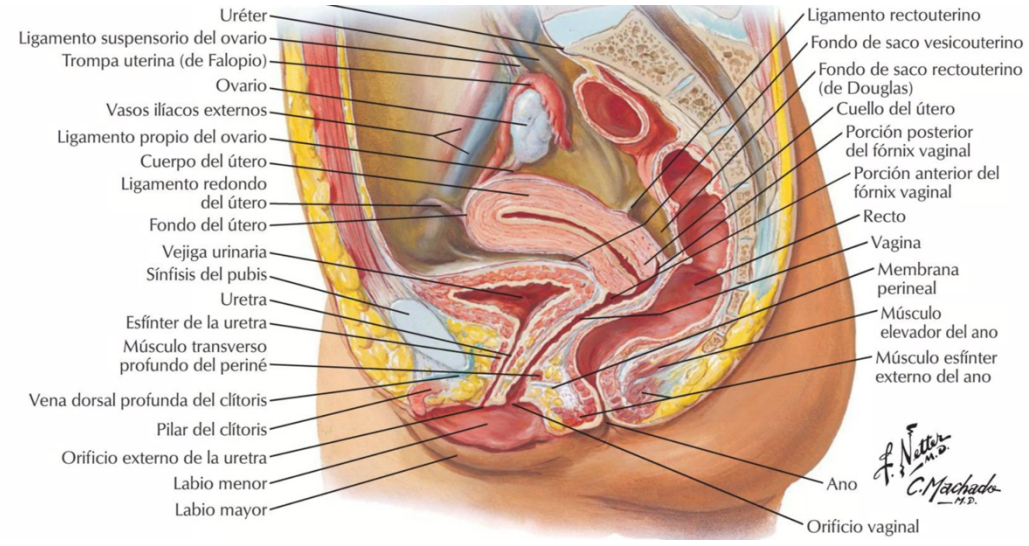


Figura 1.1. En esta imagen podemos observar las relaciones anatómicas del cérvix con el cuerpo del útero, anejos y resto de vísceras pélvicas. Imagen extraída del libro Atlas de Anatomía humana, de Frank H. Netter (6ª Edición).

El extremo inferior del cérvix es accesible a través de la vagina; que es una cavidad virtual cerrada que solo es visible mediante la exploración ginecológica a través del espéculo vaginal, conocida como colposcopia (**Figura 1.2**).

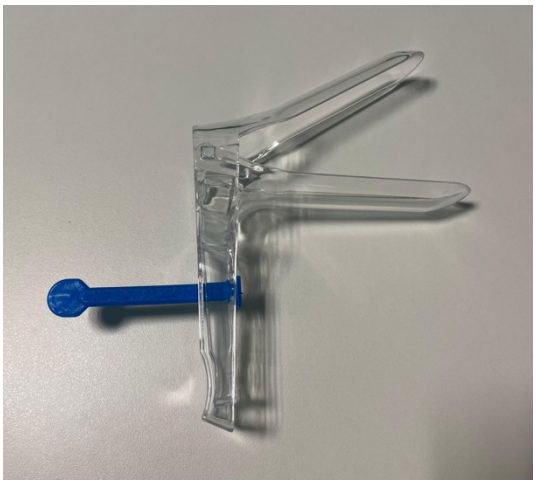


Figura 1.2. Espéculo que permite realizar una exploración ginecológica completa del cuello del útero mediante la dilatación mecánica de las paredes de la vagina, haciendo al orificio cervical externo completamente visible. Fotografía cortesía de la Dra. Ortiz González del Servicio de Oncología Radioterápica del Hospital Universitari Son Espases (Palma).

Mediante la exploración clínica, el cérvix es completamente visible y se pueden distinguir en él varias áreas muy bien diferenciadas: la cara vaginal visible del cérvix conocida como exocérvix; el orificio cervical externo (OCE) que es el punto de comunicación entre el exocérvix con el canal endocervical; y el endocérvix (Sellors & Sankaranarayanan, s. f., 2003) (**Figuras 1.3 y 1.4).** El endocérvix es distinguible clínicamente debido a que presenta un aspecto micropapilar característico, secundario a que microscópicamente contiene numerosas estructuras crípticas repletas de moco (Ross MH, Wojciech P, 2013). El orificio que comunica el canal endocervical y la cavidad uterina se conoce como orificio cervical interno (OCI) y no es visible en la exploración clínica ginecológica.

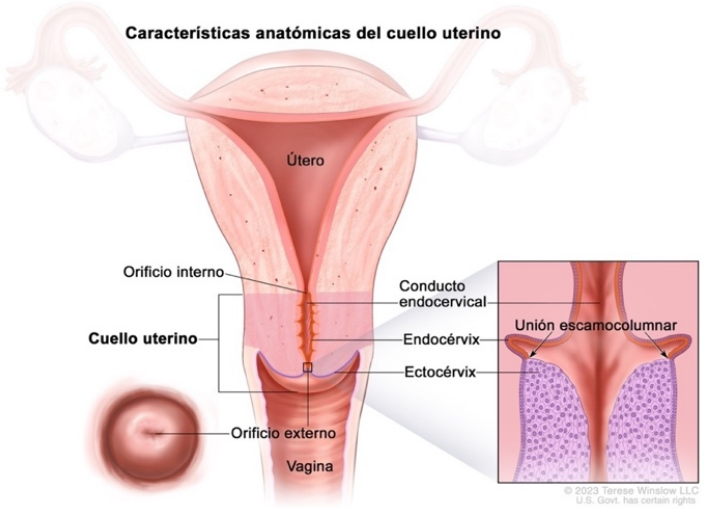


Figura 1.3. Imagen de las distintas partes del cérvix. Extraída del NCI (National Cancer Institute).

Figura 1.4. Exploración ginecológica del cuello del útero. Podemos observar al fondo la mucosa cervical y el orificio cervical externo, que da paso al canal que comunica el cérvix con el útero, conocido como canal endocervical. Fotografía cortesía de la Dra. Lara Peñaranda del Servicio de Ginecología y Obstetricia del Hospital Santa Lucía (Cartagena).

La función principal del cérvix es la de servir como canal de drenaje a las secreciones uterinas y a los productos menstruales; así como la de proteger al feto durante la gestación. Durante el embarazo actúa como una válvula que sella la cavidad uterina y la protege de agresiones e infecciones (Young et al., 2013); y durante el parto se acorta y se dilata para conducir al feto hacia la vagina (Nott et al., 2016). Actúa gracias a su forma de cuello de botella como barrera protectora de la cavidad uterina y gracias a su capacidad de producir moco, es capaz de crear el ambiente adecuado para que los espermatozoides se dirijan a las trompas de Falopio, atravesando la cavidad uterina para la fertilización.

l cérvix es una estructura dinámica que modifica sus características a lo largo del ciclo ovárico y a lo largo de la vida de la mujer (Sellors & Sankaranarayanan, s. f., 2003). Esto se debe a que el cérvix se encuentra sometido a un continuo flujo hormonal que no es constante, sino que tiene unos ciclos con niveles ascendentes y descendentes de estrógenos y progesterona, que se van repitiendo con cada ciclo y que acaban cesando una vez que se alcanza la menopausia. De esta forma y debido a la disminución del estímulo hormonal conforme avanzan los años, el cérvix va perdiendo trofismo y se va adelgazando pasando de un color rosado en mujeres pre-menopáusicas a un color blanquecino en menopáusicas.

1.1.2. HISTOLOGÍA DEL CÉRVIX UTERINO

Al examen microscópico, las distintas partes anatómicas del cérvix tienen unas características histológicas muy bien diferenciadas.

A. EL EXOCÉRVIX

El exocérvix está tapizado por un epitelio plano poliestratificado no queratinizado. Este tipo de epitelio le confiere una gran resistencia debido a que está constituido por un gran número de células planas que se sitúan en proximidad unas con otras, unidas íntimamente por uniones intercelulares conocidas como

desmosomas y por otras uniones que anclan estas células a la capa basal del epitelio conocidas como hemidesmosomas (Alazawi et al., 2003). Estas células se disponen organizadas en el epitelio en distintos estratos o capas de 15 a 20 células de grosor (Sellors & Sankaranarayanan, s. f., 2003), que son distinguibles microscópicamente, a las que la célula da nombre. De capas bajas a altas encontramos (**Figura 1.5**):

i. Estrato basal, formado por células basales y parabasales. Es el estrato más bajo, en el que las células basales son de pequeño tamaño, con un núcleo oscuro y muy escaso citoplasma (Prendiville & Sankaranarayanan, 2017), que se disponen formando una empalizada anclada a la membrana basal. Característicamente las células de este estrato se disponen perpendiculares a la superficie exocervical (Young et al., 2013). Las células parabasales son ligeramente más grandes, poligonales y se disponen en 2 capas por encima de la capa basal. En condiciones normales, en estas capas podemos encontrar mitosis ya que en esta localización las células proliferan de manera controlada para crecer, ascender, madurar y formar el resto de las capas superiores.

ii. Estrato intermedio, formado por células intermedias. Son células de gran citoplasma poligonal con núcleos pequeños que se disponen en el centro de la célula. El estrato de células intermedias es ancho, de varias capas de grosor, que se orientan paralelas a la superficie exocervical. Estas células suelen tener un color blanquecino debido a su alto contenido en glucógeno, secundario a la estimulación estrogénica (Mills, 2019). La pérdida de este contenido glucogénico puede verse; tanto en lesiones benignas inflamatorias o metaplásicas, como en cambios preneoplásicos.

iii. Estrato superficial, formado por células superficiales. Capa de 10 filas de células aplanadas, con citoplasma rosado y núcleo central muy pequeño.

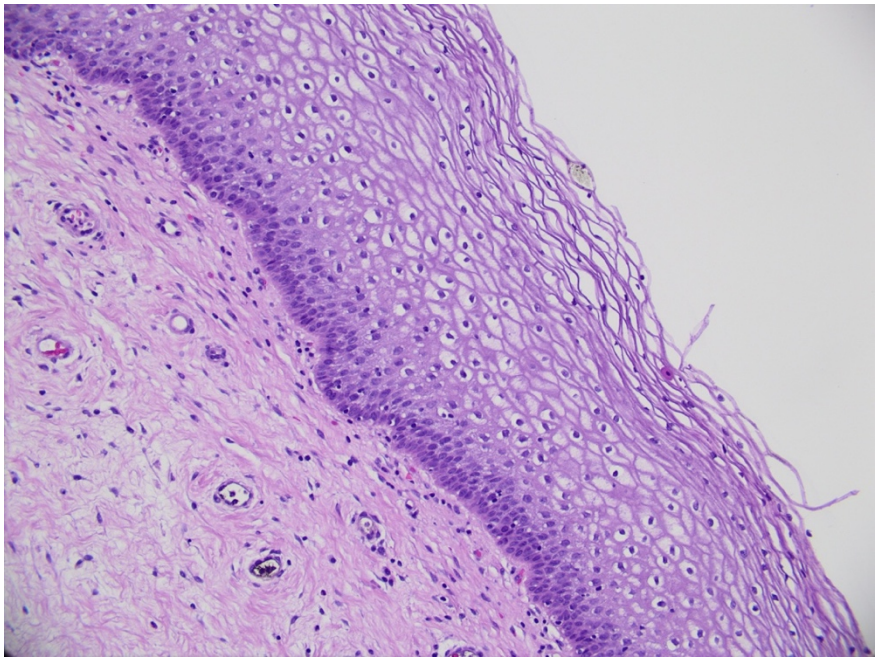


Figura 1.5. Imagen microscópica de la mucosa exocervical. Nótese los distintos estratos del epitelio escamoso que la conforman: el estrato basal, intermedio y superficial. Elaboración propia.

El dinamismo del exocérvix también es visible microscópicamente, ya que gracias al influjo hormonal las células van creciendo, madurando y ascendiendo a capas altas. La maduración hace referencia a que las células de estratos intermedios y altos se desprenden y se descaman, y la misión de las células basales y parabasales es dividirse, crecer, y restituir al resto de estratos superiores. Esta descamación permite que se libere el glucógeno del citoplasma, para que las bacterias del cérvix lo metabolicen y se produzca ácido láctico que disminuye el pH, reduciendo así el crecimiento de microorganismos (Mills, 2019). Es por ello que las células de estratos bajos son más activas mitóticamente y suelen tener un alto índice de proliferación celular, que conforme ascienden a estratos superiores disminuye hasta desaparecer. Así, en la maduración, una célula basal y parabasal sube de estratos, aumentando de tamaño del citoplasma y disminuyendo progresivamente su núcleo hasta desaparecer en capas altas.

Todo este proceso de maduración es hormono dependiente, siendo el estímulo estrogénico un paso imprescindible para que se lleve a cabo. La ausencia de estrógenos impide una maduración y glucogenación completa del epitelio, dando lugar a epitelios planos de aspecto atrófico (Prendiville & Sankaranarayanan, 2017). Esto es lo que ocurre en pacientes menopáusicas en las que ante la falta de secreción hormonal, las células no maduran más allá de la capa parabasal (Ross MH, Wojciech P, 2013), resultando en epitelios adelgazados de células basales y parabasales, que no se acompañan de células intermedias ni superficiales.

B. EL ENDOCÉRVIX

El endocérvix es un epitelio que está constituido por una única hilera de células cilíndricas que presentan mucosecreción citoplasmática. Al examen microscópico los núcleos se encuentran alineados y reposados sobre la membrana basal, quedando los citoplasmas blanquecinos cargados de moco orientados hacia la cavidad cervical. En endocérvix presenta en su tejido de sostén abundantes glándulas que drenan su secreción a la superficie (**Figura 1.6)** formando abundantes criptas, dando un aspecto micropapilar al epitelio localizado entre las criptas.

C. LA UNIÓN ESCAMOCOLUMNAR

La unión entre ambos epitelios se conoce como unión escamocolumnar (UEC), llamada así por la transición entre estos dos tipos de epitelios. Esta región coincide clínicamente con el OCE, siendo fundamental evaluarla ya que en esta región típicamente se originan las lesiones precursoras del cáncer del cérvix, secundarias a una infección por el virus del papiloma humano (del inglés, *Human Papilloma Virus* o HPV) (**Figura 1.7).**

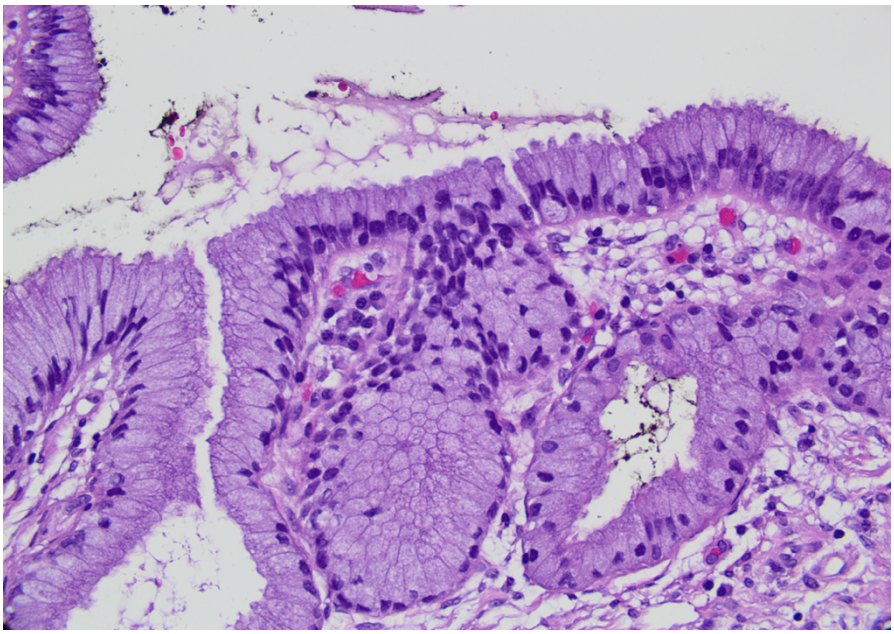


Figura 1.7. Imagen microscópica de la mucosa endocervical revestida por un epitelio cilíndrico simple mucosecretor. Elaboración propia.

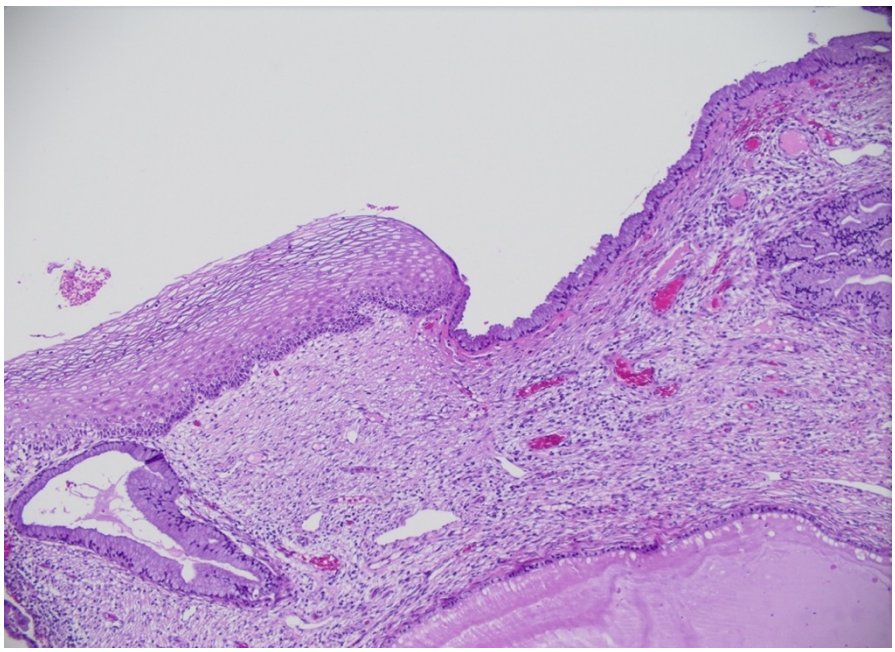


Figura 1.6. Imagen microscópica de la unión escamocolumnar en la que se observa la transición entre ambos epitelios. A la izquierda se evidencia el epitelio escamoso exocervical y a la derecha el epitelio cilíndrico simple endocervical. Elaboración propia.

1.1.3. EXPLORACIÓN CLÍNICA DEL ORIFICIO CERVICAL EXTERNO

La zona de la UEC es visible a la exploración con el espéculo, ya que esta transición entre estos dos epitelios es distinguible, al tratarse de dos epitelios de distintos grosores, que se separan a modo de "escalón" (Sellors & Sankaranarayanan, s. f., 2003). Esta división se identifica como una delgada línea divisoria entre una mucosa exocervical rosada de superficie lisa, y una mucosa endocervical rojiza de superficie micropapilar.

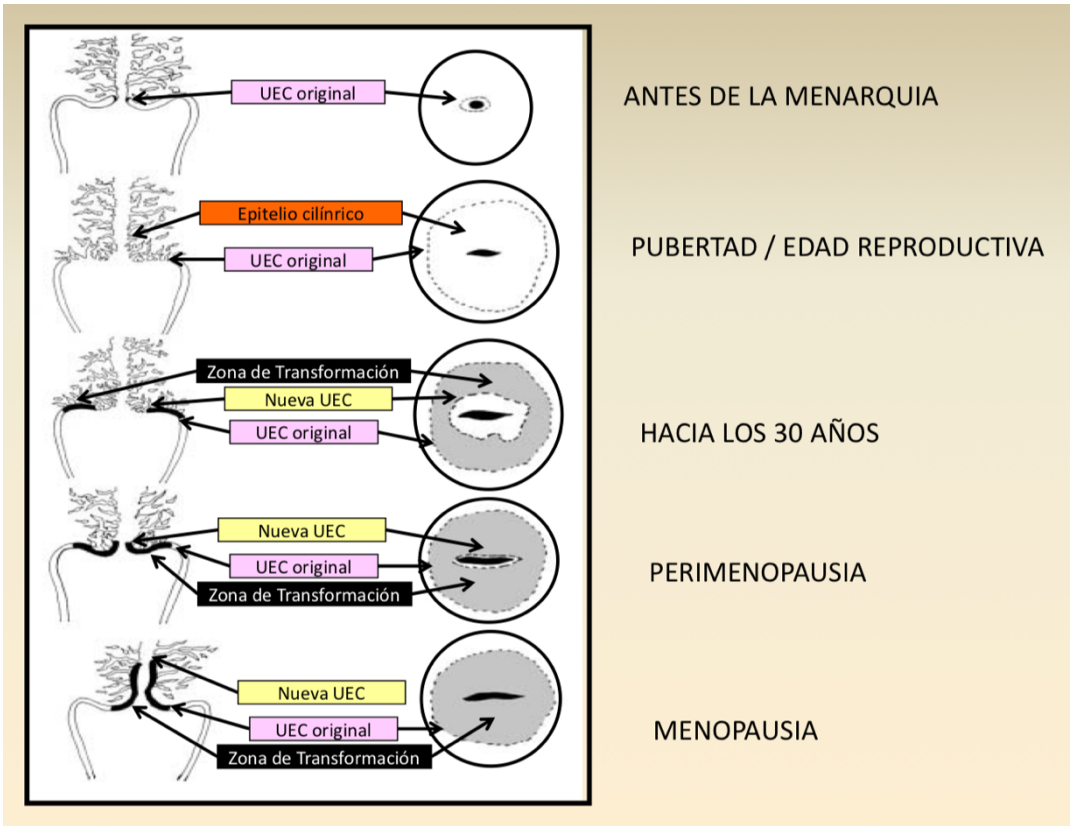


Figura 1.8. Imagen esquemática de cómo va evolucionando la unión escamocolumnar en la exploración clínica del cérvix a lo largo de los años. Imagen obtenida de "La colposcopia y el tratamiento de la neoplasia intraepitelial cervical: Manual para principiantes" de Sellors y Sankaranarayanan, modificada por Dr. Remezal Solano del Hospital Universitario Virgen de la Arrixaca (Murcia).

La localización de esa línea de separación entre las mucosas se modifica con la edad por el cambio en la influencia hormonal, con la gestación o la toma de anticonceptivos orales. Desde el nacimiento, se localiza próximo al OCE y conforme aumenta el estímulo estrogénico con la edad, esta unión entre epitelios se desplaza hacia el exocérvix, dando lugar con el tiempo a una nueva región visible de unión escamocolumnar. La región que separa la UEC original y la desplazada, se conoce como zona de transición o transformación (ZT) **(Figura 1.8**) (Sellors & Sankaranarayanan, s. f., 2003). Es fundamental localizar clínicamente la UEC y ZT ya que en estos territorios es donde asientan las lesiones precursoras del cáncer de cérvix, secundarias a la infección por HPV.

1.2. EL CÁNCER DE CÉRVIX

1.2.1. SITUACIÓN DEL CÁNCER DE CÉRVIX A NIVEL MUNDIAL

El cáncer de cérvix representa la cuarta neoplasia más frecuente en mujeres a nivel mundial y afecta a todos los países del mundo. Con el paso de los años se está observando un aumento de la incidencia de este cáncer y se están evidenciado un aumento de los números absolutos de mortalidad, según los datos publicados por GLOBOCAN (Global cancer observatory) (*Cancer Today*, s. f.). En las cifras emitidas por la Organización Mundial de la Salud (OMS) en enero de 2020, esta neoplasia presentó una incidencia de 604.127 casos nuevos y 342.000 muertes durante el año 2020 («Cervical Cancer», 2022) (**Figura 1.9**), perteneciendo las cifras más altas de nuevos casos y mortalidad a países en vías de desarrollo, sobre todo del continente africano, representando >85% de las muertes. Respecto al cáncer de cérvix avanzado, cuando es invasor, presentó una incidencia de 15 casos por cada 100.000 mujeres al año.

Representa el 3,1% de todos los diagnósticos de cáncer en el mundo, localizándose en séptimo lugar según los datos publicados en 2020, por detrás del cáncer gástrico y hepático (*Cancer Today*, s. f.). Si tenemos en cuenta únicamente los casos de cáncer que afectan a mujeres, según los datos publicados en febrero 2024

por la plataforma GLOBOCAN (*Cancer Today*, s. f.); de los casos registrados durante el año 2022, este cáncer presentó una incidencia de 662.301 casos, que representan el 6.9% de todas las neoplasias que afectan a las mujeres, frente al 6,4% que se registró en el año 2020, ocupando la cuarta posición de los cánceres que afectan a mujeres por detrás del cáncer colorrectal (**Figura 1.10**). Territorios como África oriental, África occidental, África subsahariana, África Austral, Melanesia, América del Sur y Sudeste asiático, recogen las tasas de incidencia más altas a nivel mundial (Cáncer de cérvix - SEOM: Sociedad Española de Oncología Médica © 2019, s. f.).

Respecto a la mortalidad, el cáncer de cérvix presenta una elevada tasa de mortalidad a nivel mundial, con unas cifras de mortalidad en 2020 de 341.831 casos, que representan el 7,7% de la mortalidad por cáncer en 2020 y una tasa estandarizada por edad de 13,3/100.000 habitantes (Sung et al., 2021). Las cifras de mortalidad también están en aumento ya que, según los datos recogidos por GLOBOCAN (*Cancer Today*, s. f.), en el año 2022 se alcanzaron unas cifras absolutas de mortalidad de 348.874, que representan el 8.1% de la mortalidad por cáncer en mujeres de todo el mundo y es el cuarto cáncer con mayor mortalidad en mujeres por detrás del cáncer de mama, pulmón y cáncer colorrectal (**Figura 1.11).**

Constituye la primera causa de muerte por cáncer en mujeres en varios países africanos y se estima que el 85-90% de las muertes por cáncer de cérvix tienen lugar en países subdesarrollados, que presentan una tasa de mortalidad 18 veces superior a los países desarrollados (**Figura 1.12).**

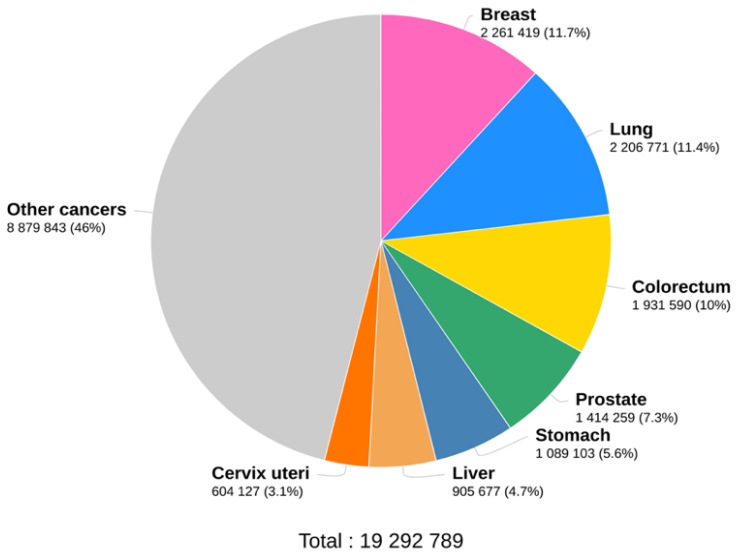


Figura 1.9. Casos de cáncer a nivel mundial en ambos sexos registrados en el año 2020. Gráfico obtenido de GLOBOCAN.

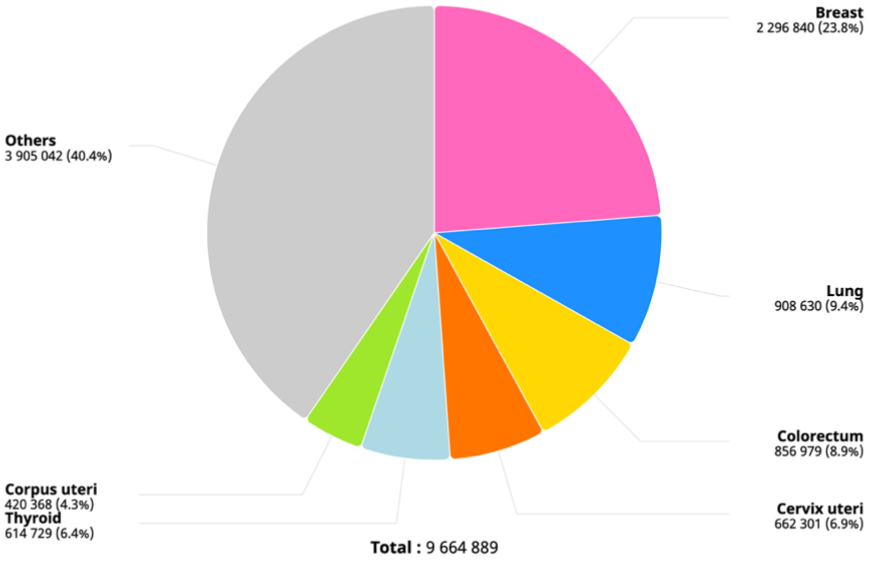


Figura 1.10. Casos de cáncer a nivel mundial en mujeres registrados en el año 2022. Gráfico obtenido de GLOBOCAN.

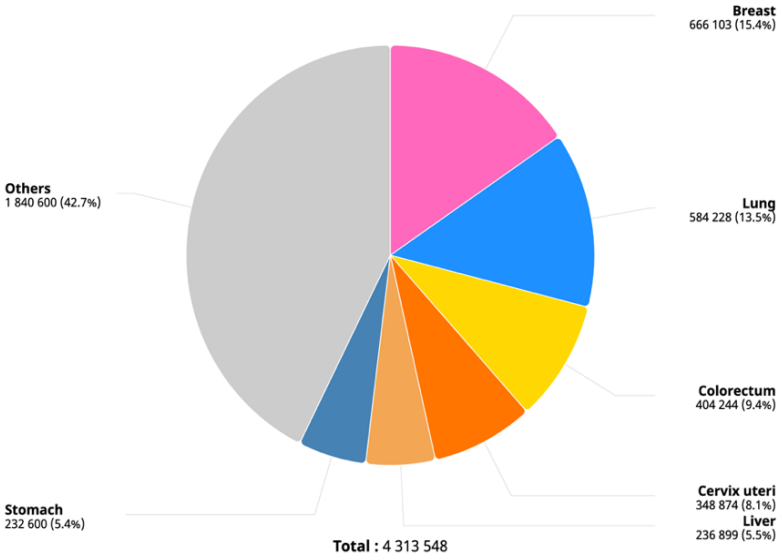


Figura 1.12. Números absolutos de mortalidad atribuidas al cáncer en mujeres durante el año 2022. El cáncer de cérvix ocupa la cuarta posición con un 8,1% de los fallecimientos. Gráfico obtenido de GLOBOCAN.

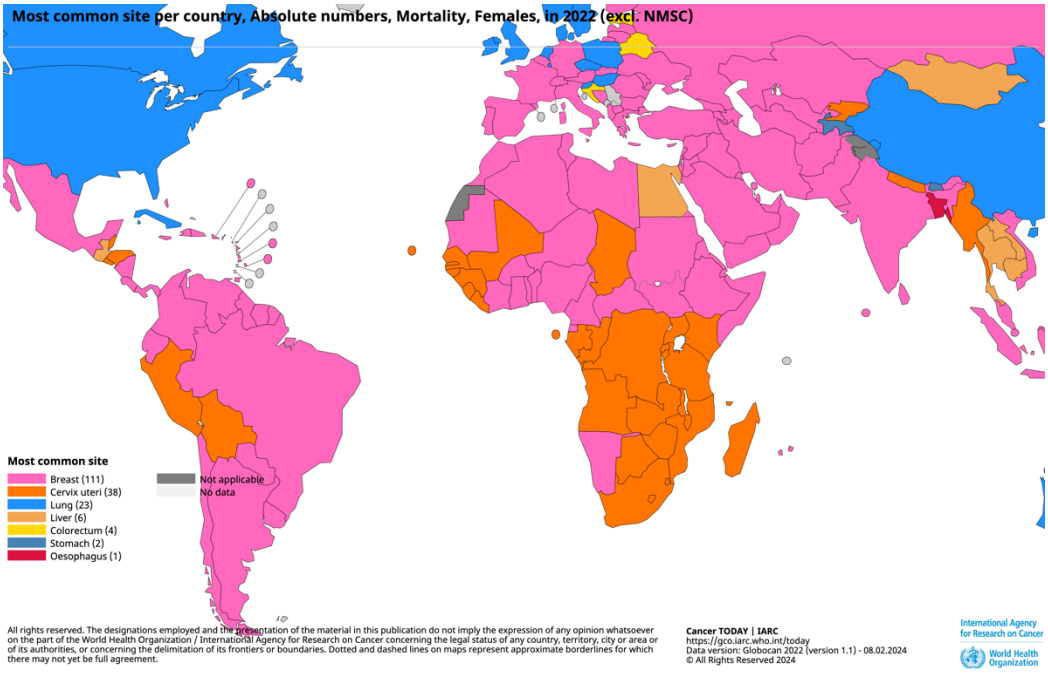


Figura 1.11. Datos de mortalidad por cáncer en mujeres en el 2022, donde se observan una mayor distribución de los fallecimientos por cáncer de cérvix en países de África y América del Sur. Mapamundi obtenido de GLOBOCAN.

Afortunadamente, el cáncer de cérvix es una de las neoplasias que mejor responde al tratamiento si se proporciona un tratamiento precoz. El conocimiento de la historia de la enfermedad y de sus factores desencadenantes; así como el desarrollo de métodos diagnósticos y de tratamiento, han establecido las bases de prevención del cáncer. Los programas de cribado tratan de buscar a las pacientes que presentan estadios subclínicos de la enfermedad; es decir, que están sanas pero presentan riesgo de desarrollar cáncer de cérvix por presentar lesiones premalignas (Hakama & Räsänen-Virtanen, 1976). En los países más desarrollados se han implantado estrategias de cribado y vacunación contra la infección por el HPV, que están dando buenos resultados, ya que están permitiendo que se pueda controlar mejor la enfermedad, gracias a la realización de un tratamiento precoz («Cervical Cancer», 2022).

Sin embargo, en los países en vías de desarrollo; tanto el acceso a estas medidas de prevención, como a los distintos tratamientos (cirugía, quimioterapia o radioterapia) es muy limitado, por lo que empeora mucho el pronóstico y eleva las cifras de mortalidad. Estas diferencias socioeconómicas, sociales y de servicios sanitarios, hacen que la enfermedad se diagnostique en estadios avanzados, cuando la enfermedad da clínica (como sangrado genital y dolor) elevando las cifras de incidencia y mortalidad, que difieren mucho respecto a las observadas en países más desarrollados. En 2020, un 90% de los casos nuevos de cáncer de cérvix y las muertes asociadas al mismo registrados a nivel mundial, se desarrollaron en países con ingresos medios-bajos (Sung et al., 2021).

1.2.2. SITUACIÓN DEL CÁNCER DE CÉRVIX EN ESPAÑA

La Sociedad Española de Oncología Médica (SEOM) (*Cáncer de cervix - SEOM: Sociedad Española de Oncología Médica © 2019*, s. f.), facilita datos sobre la incidencia y prevalencia del cáncer en nuestro país. Según los datos recogidos durante el año 2022 en España, el cáncer de cérvix presentó una incidencia en mujeres de 5.4 casos/100.000 mujeres/año y una mortalidad de 1.6 casos/100.000 mujeres/año. Se trata del onceavo cáncer con más incidencia en mujeres en España (**Figura 1.13).** Según las estadísticas de GLOBOCAN, se está observando un descenso en su incidencia al comparar los datos con las cifras registradas en el 2020 de 10.8 casos/100.000 mujeres/año y una mortalidad de 3.6 casos/100.000 mujeres/año.

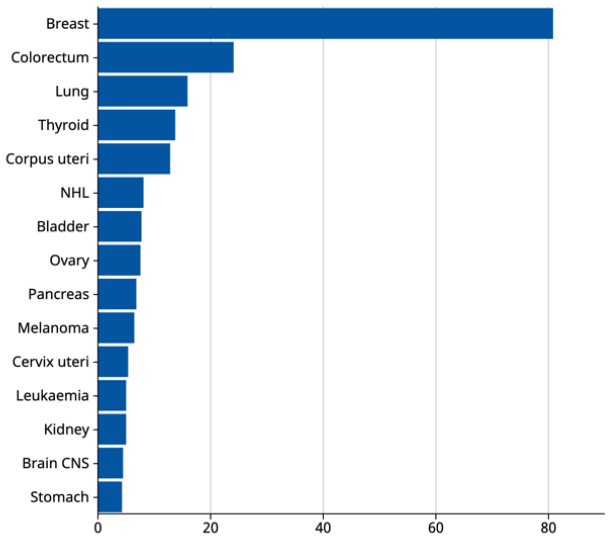
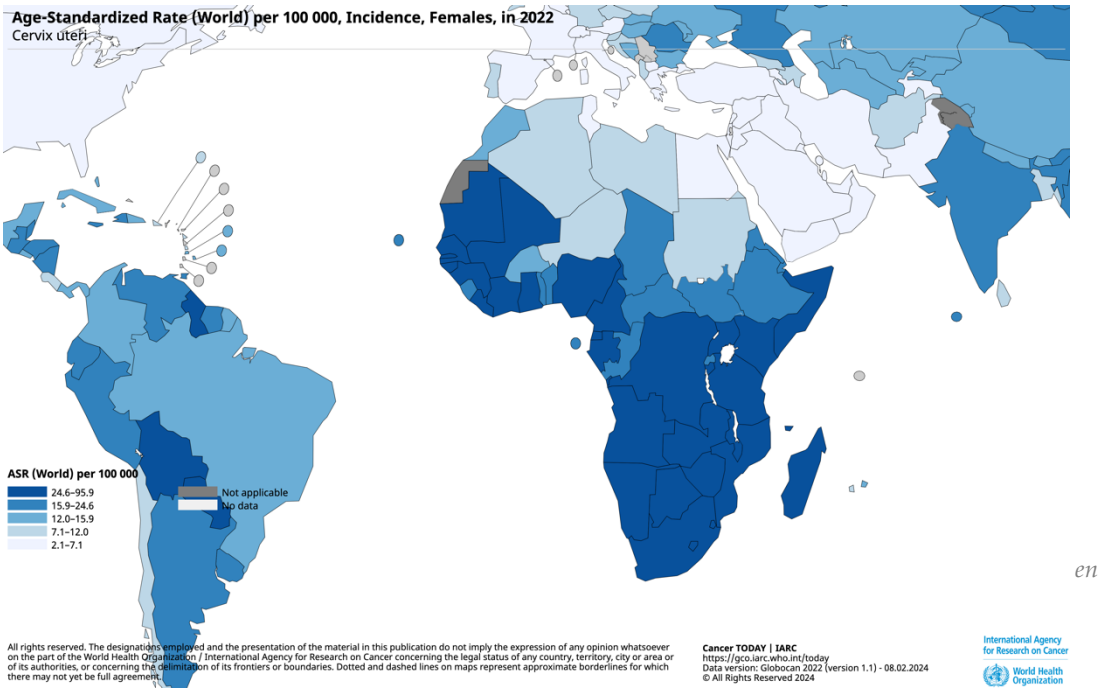


Figura 1.13. Incidencia en España de distintos tipos de cáncer por cada 100.000 mujeres / año, según los datos recogidos en GLOBOCAN en 2022. Gráfico obtenido de GLOBOCAN.

A pesar del aparente descenso en su incidencia, el cáncer de cérvix ocupa un lugar importante en el diagnóstico y tratamiento del cáncer, por implicar grandes costos sociosanitarios y acarrear una gran morbilidad si no se trata a tiempo. En el año 2020, se registraron 606.084 casos de cáncer de cérvix en nuestro país, que representaron el 6,5% de todos los diagnósticos de cáncer.

Al comparar la situación del cáncer de cérvix en España con el resto de países, nuestro país tiene de las incidencias más bajas en Europa, junto a Holanda, Francia, Italia y Grecia (*Libro Blanco Anat. Patologica-2013*, s. f.); y de las mortalidades más bajas junto a Holanda, Suiza,, Dinamarca, Italia y Grecia (**Figura 1.14**). Según Álvarez y cols. (Álvarez García et al., 2024), en España ocurren 2000 casos de cáncer de cérvix al año, que implican 700-800 muertes anuales, con un promedio mínimo de dos muertes diarias.



en

Figura 1.14. Incidencia por cada 100.000 mujeres del cáncer de cérvix recogidos en el año 2022 en todo el mundo. Mapamundi obtenido de GLOBOCAN.

1.2.3. ANTECEDENTES Y ESTUDIO COLPOSCÓPICO DEL CÁNCER DE CÉRVIX

El cérvix puede albergar distintos tipos de cánceres dependiendo de a partir de qué célula se desarrolla el mismo. A partir de una célula se puede producir una proliferación descontrolada que, junto a factores ambientales y mutacionales, desarrollará el cáncer. El cáncer más frecuente del cérvix, es aquel que se origina de las células escamosas del epitelio plano poliestratificado exocervical, conocido como carcinoma escamoso o de células escamosas, nomenclatura que hace referencia a su procedencia. Las primeras descripciones de este cáncer comenzaron en 1886 por Williams (Kolstad, 1970) y en 1900 por Cullen (Younge, 1965), que describieron la histología del carcinoma de células escamosas in situ.

Para la localización de las lesiones neoplásicas o preneoplásicas a la exploración física del cérvix mediante colposcopia, hay descritas algunas técnicas, como el test de Schiller o la técnica del ácido acético. Se tratan de una técnicas colorimétricas que marcan de distintos colores la mucosa sana y las zonas preneoplásicas, facilitándole al ginecólogo la biopsia dirigida de las lesiones sospechosas de malignidad para su confirmación histopatológica **Figura 1.15**).

Figura 1.15. Colposcopia con test de Schiller positivo, con el que se evidencia un área color mostaza pobre en glucógeno sospechosa de displasia de la mucosa cervical normal rica en glucógeno marcada de marrón. Fotografía cortesía de la Dra. Lara Peñaranda del Servicio de Ginecología y Obstetricia del Hospital Santa Lucía (Cartagena).

1.2.4. NOMENCLATURA DE LA DISPLASIA CERVICAL: TERMINOLOGÍA LAST

Las lesiones precursoras del cáncer de cérvix, también conocidas como displásicas o preneoplásicas, son aquellas que llevadas a una libre evolución tienen riesgo de progresión al cáncer de cérvix. Por lo tanto, este tipo de entidades hacen referencia a epitelios exocervicales que presentan alteraciones citológicas y arquitecturales previas al desarrollo del cáncer. La incidencia de lesiones displásicas en pacientes detectadas por cribado citológico se sitúa entre el 1,1 y el 3,7% según Anttila y cols. (Anttila et al., 2010). Según Castillo y cols. (Castillo et al., 2016) e Ibáñez y cols. el 60-70% de las pacientes con cáncer cervical en España no se habían realizado ninguna prueba de screening para detectar lesiones precursoras durante los 10 años previos al diagnóstico.

El campo de la patología cervical ha experimentado varios cambios en su nomenclatura con el paso de los años con el objetivo de unificar y simplificar conceptos para facilitar la comunicación entre profesionales. En 1949 Papanicolau comienza a designar a estos cambios preneoplásicos en la biopsia como "displasia" y en la citología como "discariosis". Posteriormente, en 1966 Richart (Martínez, 2006) establece su clasificación de "neoplasia intraepitelial cervical" (acrónimo CIN) en la que reconocía 3 grados de severidad de la displasia de acuerdo a la arquitectura del tejido, conocidos como CIN 1, 2 y 3. A pesar de que este sistema fue ampliamente aceptado durante más de 20 años, se observó una baja seguridad en el diagnóstico de las displasias menos severas, creando confusión entre los profesionales.

Para favorecer la comunicación entre especialistas y la buena correlación entre los resultados objetivados en biopsia y citología, y tras los nuevos conocimientos en la carcinogénesis cervical, la OMS decidió en 1970 unificar la terminología de la displasia y simplificarla en una clasificación binaria, empleando los términos "lesión intraepitelial escamosa" (*squamous intraepitelial lesion en inglés,* con el acrónimo SIL) que incluye solo 2 grados de displasia: displasia de bajo grado (o LSIL, del inglés, *low grade squamous intraepitelial lesion*) y de alto grado (o HSIL,

del inglés *high grade squamous intraepitelial lesion*). En esta nueva clasificación LSIL engloba el CIN1; y HSIL el CIN2, CIN3 y carcinoma in situ de Richart. A raíz de esta unificación, entre 1988-1991 el Instituto Nacional del Cáncer de Estados Unidos se reunió en Bethesda (Maryland, Estados Unidos) para establecer una nomenclatura para los informes citológicos de las alteraciones celulares resultantes de la infección por el HPV (Nayar & Wilbur, 2015), que dio lugar al *Sistema Bethesda de diagnóstico de la citología cervicovaginal*, que tras actualizaciones en 1991, 2001 y 2014 está implantado hoy día en numerosos laboratorios por todo el mundo.

La asociación entre displasia e infección por HPV no es exclusiva del cérvix y se observa en otros territorios como vagina, vulva, ano, región perianal o pene (Darragh et al., 2013). En 2013 se publicaron unas recomendaciones del Colegio Americano de Patólogos y de la Sociedad Americana de Colposcopia y Patología Cervical sobre generalizar la nomenclatura de "lesión intraepitelial escamosa" a todo estos territorios del tracto anogenital que compartían esta relación con la infección por HPV, conocida como *terminología LAST (Low Anogenital Squamous Terminology*) (Darragh et al., 2013).

Hoy día, empleamos la terminología de "lesión intraepitelial escamosa" para el diagnóstico de displasias en citología, mientras que también se añade la terminología de Richart para la biopsia, debido a que hay publicados estudios que demuestran distintos comportamientos de las lesiones CIN 2 y CIN 3 a pesar de que se engloben ambas como un HSIL (McCredie et al., 2008), (Moscicki et al., 2010).

Tabla 1.1. Tabla comparativa con la evolución de la nomenclatura de la displasia cervical con los años y cuáles son las equivalencias entre las distintas clasificaciones. Elaboración propia.

Displasia / CIS Años 49-69 → 4 grados	Displasia leve	Displasia moderada	Displasia severa	Carcinoma in situ
CIN *Richart* (años 66-88) → 3 grados	CIN 1	CIN 2	CIN 3	
SIL *Bethesda* (desde 1988) / *Terminología LAST* (2013) → 2 grados	Displasia de bajo grado LSIL	Displasia de alto grado HSIL		

1.2.5. HISTOPATOLOGÍA DE LAS LESIONES PRECURSORAS DEL CÁNCER DE CÉRVIX

Es importante conocer los cambios celulares observados en las displasias cervicales cuando se estudian con una biopsia, ya que la citología no es una técnica diagnóstica definitiva y generalmente va a requerir una confirmación histopatológica que marcará el tratamiento de la paciente. En términos generales, las pacientes que obtienen un diagnóstico de HSIL en citología se remiten para estudio por colposcopia para la toma de biopsia y estudio histopatológico, mientras que los ASC-US y LSIL se controlan con citologías periódicas (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013).

En condiciones normales, un epitelio exocervical normal presenta una similitud en la forma y tamaño de las células que lo componen, según su distribución ordenada por estratos en el epitelio, dando una apariencia homogénea de todas las capas. Es decir, las células de capas bajas son similares entre sí, al igual que las células de capas intermedias y superficiales.

Una alteración en esta homogeneidad celular nos pone en alerta ante la presencia de displasia previa al desarrollo de cáncer y se traduce como distintos cambios nucleares que son evidentes tanto en citologías como en biopsias. Estos cambios pueden combinarse entre sí y son, de manera general:

- Aumento de la relación núcleo-citoplasma. Consiste en un aumento del tamaño del núcleo respecto al área del citoplasma. Este hallazgo es prácticamente constante en todas las lesiones displásicas; tanto en citología como en biopsia.

- Pérdida de la polaridad nuclear. Hace referencia a la pérdida de una disposición central del núcleo dentro del citoplasma. En condiciones normales, los núcleos están en el centro de la célula, localizándose centrados en el citoplasma. En las lesiones displásicas esta polaridad se pierde,

encontrándose los núcleos en distintos puntos del citoplasma, dando una apariencia desordenada.

- Pleomorfismo nuclear. Hace referencia a la variedad en la forma y tamaño de los núcleos. Dentro del mismo estrato del epitelio, lo esperable es que todos los núcleos tengan una forma y tamaño semejante, ya que se encuentran en el mismo estadio madurativo. El pleomorfismo indica esa falta de homogenización y la presencia de una gran variedad en los núcleos dentro de los mismos estratos.

- Cromatina en grumos. Consiste en la imagen que da la cromatina cuando se condensa de manera irregular en el núcleo. A diferencia de esta, la cromatina de los núcleos de células sanas en fina y granular.

- Mitosis en capas altas del epitelio. Hacen referencia al aumento del índice proliferativo de células fuera de su territorio habitual, que es la capa basal y parabasal.

- Nucléolos. En condiciones normales las células epiteliales al examen citológico no tienen nucléolos. La presencia de nucléolos deben alertar de la presencia de lesiones displásicas severas o de un carcinoma infiltrante.

Existen otro tipo de cambios nucleares que pueden aparecen en las células que han sido infectadas por el HPV, condición necesaria para el desarrollo de displasias y carcinomas de algunos territorios, como la mucosa cervical o anal. Estos hallazgos pueden coexistir con las alteraciones citoarquitecturales displásicas definidas anteriormente o manifestarse de manera aislada en células infectadas que no han llegado a desarrollar cambios displásicos. Estas alteraciones son:

- Presencia de células disqueratósicas. Consisten en células escamosas diminutas con citoplasma eosinófilo o anaranjado denso con núcleos pequeños e hipercromáticos.

- Células coilocíticas o coilocitos. Son células de citoplasma amplio con núcleos irregulares e hipercromáticos que característicamente presentan un halo blanquecino perinuclear. Esta característica es típica de la displasia de bajo grado o LSIL, en la que se produce la síntesis de una cápside proteica del HPV que manifiesta sus características citopáticas en las capas altas del epitelio, dando este aspecto celular tan característico.

- Células binucleadas o multinucleadas. Consiste en la presencia de dos o más núcleos en una célula. Este cambio no es exclusivo de la infección por HPV, sino que también pueden aparecer en otros procesos infecciosos virales, como infecciones por el virus herpes simple.

En las biopsias de cérvix, la distribución de estos cambios celulares por el epitelio cervical determina el grado de displasia de las lesiones precancerosas.

Siguiendo la terminología de Richart, cuando estas alteraciones alcanzan el tercio inferior del epitelio lo conocemos como CIN 1, cuando alcanzan tercio medio como CIN 2 y tercio superior CIN 3. De acuerdo a la terminología LAST y al Sistema Bethesda de diagnóstico de la citología cervicovaginal; CIN 1 se corresponde con LSIL y CIN 2 y 3 con un HSIL **(Figuras 1.16, 1.17 y 1.18).**

Existen técnicas de inmunohistoquímica que pueden ayudar a la hora de diferenciar si una lesión escamosa es reactiva o displásica, herramienta útil cuando se observan cambios reactivos en el epitelio que simulan displasias cervicales; como la metaplasia escamosa inmadura, la atrofia o los cambios reparativos. Se trata de una técnica de laboratorio que permite detectar un determinado antígeno en el tejido cuando se revela colorimétricamente tras su adhesión con su anticuerpo. La molécula p16 interviene en el ciclo celular y está sobreexpresada en las lesiones epiteliales causadas por el HPV, siendo muy específica en lesiones displásicas de alto grado causadas por virus de alto riesgo (Bergeron et al., 2010), (Ordi et al., 2009), (Stoler et al., 2018). Otra técnica muy útil es el ki67 que revela las células proliferativas del epitelio que estará sobreexpresado más allá de la capa parabasal en las displasias y neoplasias cervicales (**Figura 1.19).**

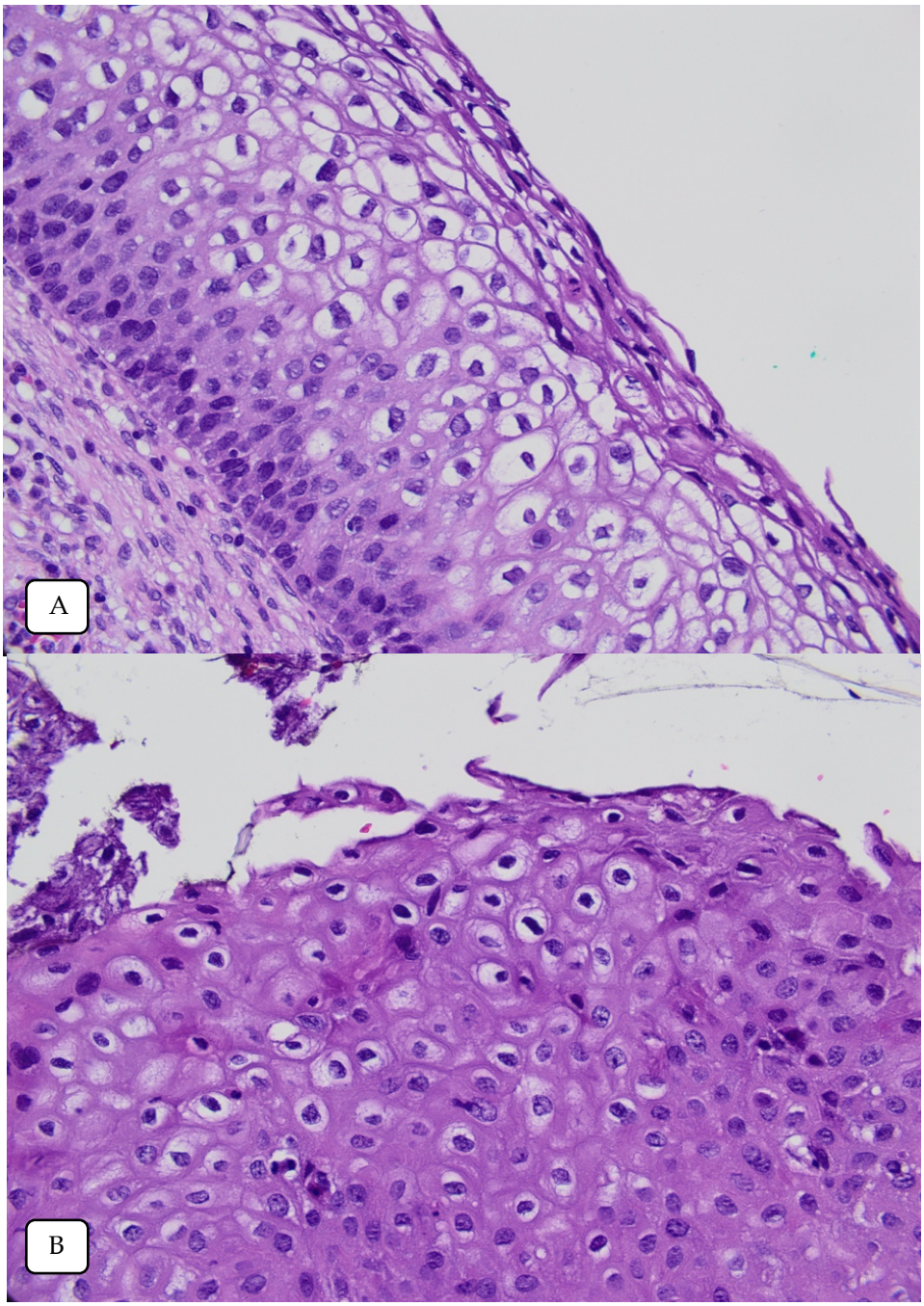


Figura 1.16. Imagen microscópica de una lesión intraepitelial escamosa de bajo grado (LSIL). Nótese la presencia de alteraciones citoarquitecturales en capas bajas del epitelio (arriba) y la presencia de coilocitos en capas altas del epitelio (abajo). Elaboración propia.

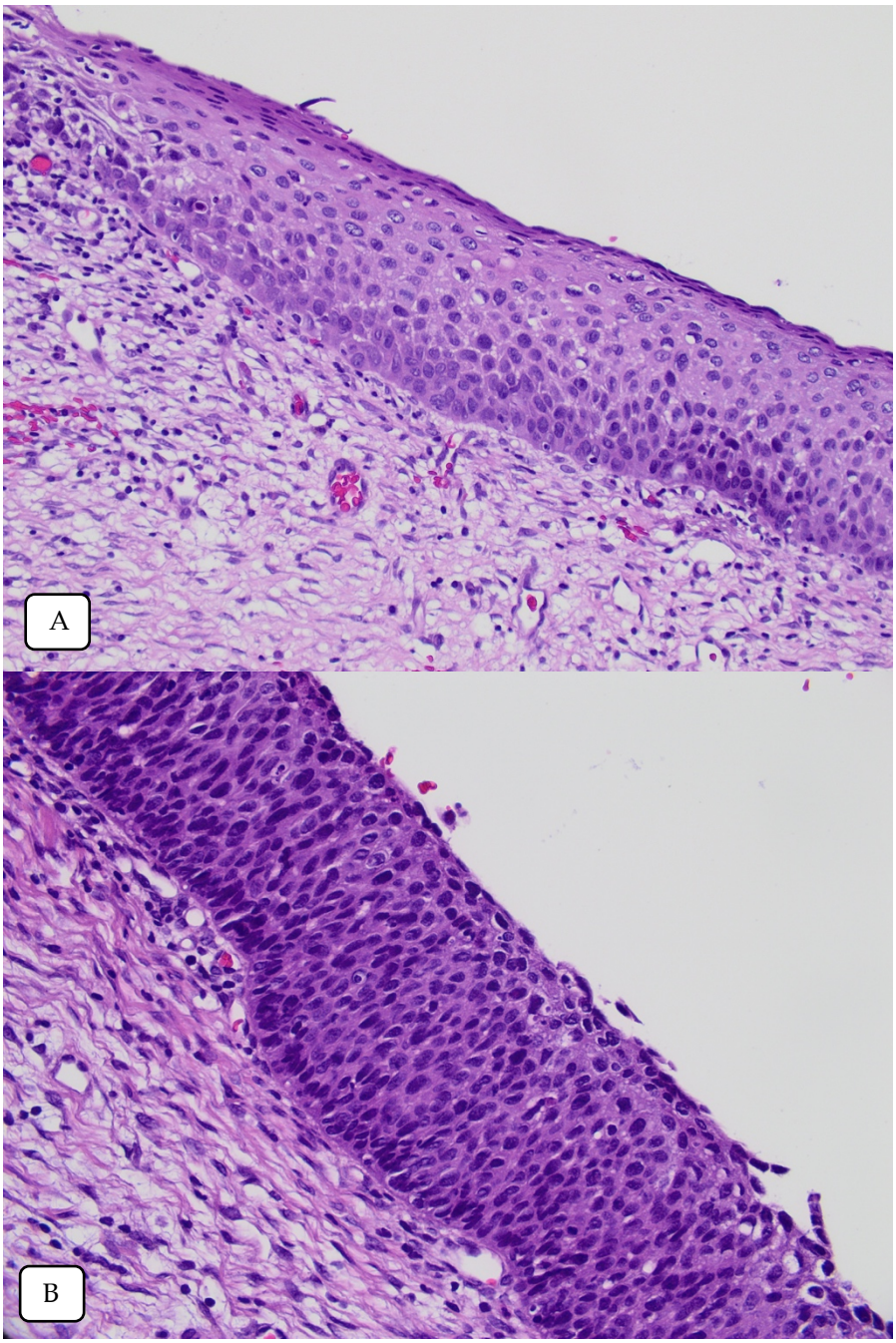


Figura 1.17. Imagen microscópica de una lesion intraepitelial escamosa de alto grado HSIL. Observamos que las células presentan alteraciones citoarquitecturales que ocupan estratos intermedios del epitelio compatibles con un CIN 2 (arriba) y superficiales compatibles con CIN 3 (abajo). Elaboración propia.

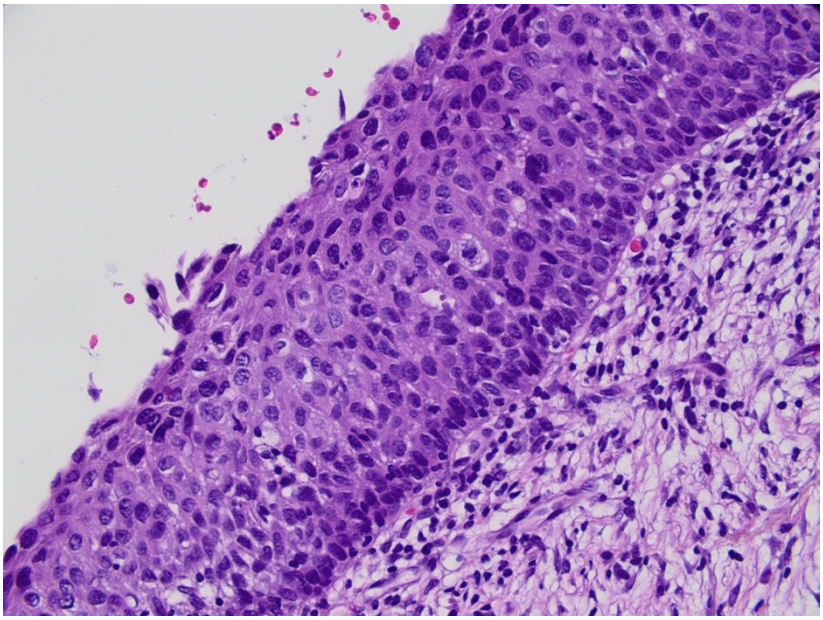


Figura 1.18. Detalle del HSIL por CIN 3. Nótese la intensa alteración citoarquitectural que ocupa por completo el epitelio con alteración de la polaridad nuclear, aumento de la relación núcleo-citoplasma, presencia de pleomorfismo nuclear y mitosis. Elaboración propia.

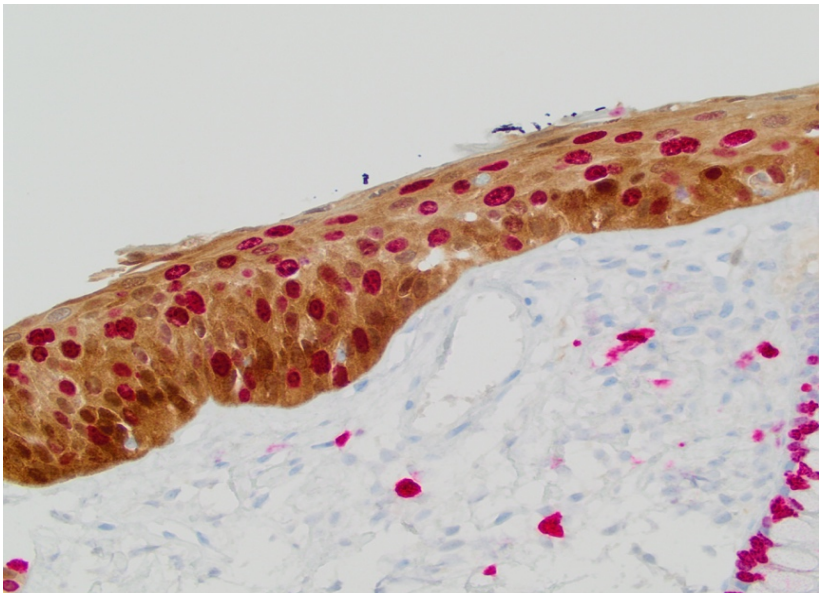


Figura 1.19. Doble tinción inmunohistoquímica con p16/ki67. Esta tinción marca de color marrón el p16 que señala el citoplasma de las células infectadas por el HPV y de color rojo el núcleo de células proliferativas, con que muestra positividad más allá de la capa parabasal. Podemos interpretar que estas células están infectadas y que presentan cambios displásicos que llegan a capas altas del epitelio, compatible con un HSIL por CIN 3. Elaboración propia.

1.2.6. ESTRATEGIA DE CRIBADO DEL CÁNCER DE CÉRVIX

El cribado o screening es un servicio publico que se emplea para detectar aquellas personas en riesgo de presentar una enfermedad, con el fin de proporcionar tratamientos iniciales que produzcan una menor morbimortalidad. En general, un cribado debe cumplir unas pautas para que aporte beneficios a la población, entre las que destaca que en la historia natural del cáncer exista un estadio previo precanceroso que se pueda detectar de forma temprana y disponer de alguna prueba de cribado que sea capaz de detectar este estadio preneoplásico (Torné A. et al., 2022).

Aplicado al cáncer de cérvix, los programas de cribado consisten en la detección de aquellas mujeres sanas que presentan riesgo de presentar cáncer, buscando lesiones premalignas o displásicas que aparecen en estadios subclínicos; tanto mediante la citología cervicovaginal (buscando la detección de lesiones intraepiteliales escamosas) como con la determinación del HPV por una reacción en cadena de la polimerasa (PCR) (Hakama & Räsänen-Virtanen, 1976).

El principal objetivo del cribado es reducir la mortalidad (Torné A. et al., 2022) y gracias al mismo aumentar la supervivencia al 92% en 5 años, según los estudios de Saslow y cols. (Saslow et al., 2012). Con una cobertura del 70-80% de la población mantenida durante años, se puede alcanzar una disminución de la mortalidad del 80% (Morales Martínez et al., 2015).

Sin embargo, el cribado está sujeto a una serie de riesgos de error que deben ser tomados en cuenta (Arbyn et al., 2010) como: la presencia de falsos positivos que obligan a realizar más pruebas complementarias para confirmación; falsos negativos que retrasan el diagnóstico y tratamiento; el sobrediagnóstico de lesiones cervicales de bajo grado que dejadas a su evolución nunca habrían progresado al cáncer; y el sobretratamiento de lesiones que dejadas a su evolución habrían regresado espontáneamente sin necesidad de tratamiento.

El cribado del cáncer de cérvix se implantó en Europa en Finlandia, Luxemburgo y Suecia en los años 60. Puede ser de dos tipos:

- Cribado oportunista. Se aplica únicamente a las mujeres que acceden al sistema sanitario por algún otro motivo de salud. Las mujeres excluidas suelen ser aquellas de menor cultura sanitaria, menos recursos sanitarios o de edad avanzada. Este cribado es limitado ya que no garantiza una cobertura universal de la población. En España, en general el cribado es oportunista.

- Cribado poblacional. Es un cribado universal y equitativo que se aplica a un conjunto de mujeres seleccionadas como grupo diana por cumplir unas condiciones determinadas. Debido a su alto coste económico solo se aplica a algunas comunidades autónomas de nuestro país como Castilla León, La Rioja y Castilla la Mancha (Torné Blade et al., 2014).

Hasta el desarrollo de la determinación de la infección de HPV por PCR, el cribado del cáncer de cérvix se realizaba exclusivamente a través de la citología cervicovaginal en la que se ha fundamentado el cribado durante más de 50 años (Torné A. et al., 2022). De manera global en la práctica, la citología cervicovaginal es el método más utilizado de cribado a nivel mundial aunque la elección de la técnica depende de la educación sanitaria del país y sus recursos sociosanitarios.

Se ha demostrado en los últimos años que un cribado primario con la determinación de HPV por PCR tiene una mayor sensibilidad y presenta un mejor rendimiento que la citología para la detección de lesiones precancerosas (Torné A. et al., 2022). Como inconveniente de la técnica, destaca que la detección de la infección por HPV es una técnica más costosa y menos específica que la citología. Esta diferencia en la especificidad se justifica porque el estado infectivo por HPV no implica necesariamente el desarrollo de displasia o neoplasia, ya que en un porcentaje no desdeñable de las pacientes, la infección se resuelve sola sin producir lesiones orgánicas. Por otro lado, la infección por el virus es una causa sin la cual no se puede desarrollar este tipo de displasias y neoplasias, por lo que la detección

de la infección permite seleccionar con certeza las pacientes que presentan riesgo de desarrollar cáncer de cérvix.

En la actualidad la guía de la Sociedad Española de Patología Cervical y Colposcopia (AEPCC), publicada en 2022, establece que el cribado se debe realizar de la siguiente forma (Torné A. et al., 2022) :

a. En pacientes no vacunadas: realización de una citología entre 25 y 35 años (en intervalos de 3 años) y determinación de HPV entre 35 y 65-70 años (en intervalos de 5 años).

b. En pacientes vacunadas: iniciar directamente con prueba HPV entre 30-65 años (en intervalos de 5 años).

De acuerdo a las recomendaciones de la guía, la determinación del virus se realiza a partir de los 30-35 años según el estado de vacunación de la paciente. En mujeres más jóvenes no se recomienda la determinación del HPV porque puede dar lugar a un sobrediagnóstico, ya que esta infección es muy prevalente en mujeres menores de 30 años, comportándose como infecciones transitorias que no producen patología y son eliminadas por el organismo. En este rango de edad, la AEPCC sí recomienda realizar una citología cervicovaginal ya que presenta una mayor especificidad.

En España el 60-70% de las mujeres con cáncer de cérvix no presentan ningún estudio de cribado durante los 10 años previos (Ibáñez et al., 2015). Por lo tanto, España está en una etapa de transición hacia el cribado poblacional con la determinación de HPV en mujeres de más de 30 años, aunque a día de hoy varía entre las distintas comunidades autónomas; según la Orden SCB/480/2019 que modifica el Real Decreto 1030/2006 de la cartera de servicios comunes del Sistema Nacional de Salud de 2019.

1.3. VIRUS DEL PAPILOMA HUMANO Y SU INFLUENCIA SOBRE EL CÁNCER DE CÉRVIX

1.3.1. GENOTIPOS DEL VIRUS DEL PAPILOMA HUMANO

Con el avance de los años se ha demostrado la capacidad carcinogénica de algunos virus en el desarrollo de determinadas neoplasias del organismo. En 1974 Hausen detectó ADN del virus del papiloma humano en las biopsias de cáncer cervical, por lo que relacionó la infección del virus con el desarrollo del cáncer, por lo que fue galardonado en 2008 con el Premio Nobel de la Medicina, según Cuestas (Cuestas E., 2008).

El virus del papiloma humano tiene la capacidad de integrar su ADN viral en el ADN de la célula huésped, provocar una alteración en el genoma y sintetizar oncoproteínas que alteran los mecanismos de apoptosis y reparación celular, favoreciendo una división celular descontrolada y el desarrollo del cáncer. El virus que infecta el tracto genital femenino se conoce como *Alphapapilomavirus* y su forma de transmisión más frecuente es a través del contacto sexual de las mucosas vaginal y cervical a través de la penetración o sexo oral.

Podemos encontrar distintos genotipos HPV con distinto riesgo carcinogénico:

- HPV de bajo riesgo que causan lesiones de poca agresividad biológica como lesiones intraepiteliales escamosas de bajo grado (LSIL/CIN1), las verrugas genitales o condilomas, que por lo general no suelen malignizarse (Bernard et al., 2010). A este grupo pertenecen los genotipos 6, 11; entre otros.

- HPV de alto riesgo que causan lesiones premalignas de gran agresividad biológica con gran riesgo de progresar a cáncer, como el HSIL. Pertenecen a este grupo los genotipos 16 y 18 que producen el 70% de los cánceres de

cérvix según Torné y cols. y otros genotipos con poder carcinogénico menor como el 31, 33, 35, 39, 45, 51, 52, 56, 58 y 59 (Torné Blade et al., 2014).

1.3.2. EPIDEMIOLOGÍA

Se estima que aproximadamente el 75% de las mujeres y el 85% de los hombres desarrollarán la infección en algún momento de su vida (Hartwig et al., 2015), (Alemany L. et al., 2016). La aparición de la infección se relaciona directamente con el inicio de la actividad sexual, por lo que la mayor prevalencia se produce en mujeres menores de 30 años (25-30%), y posteriormente se estabiliza a partir de los 40 años (Castellsagué et al., 2012). Sin embargo, en la mayoría de los casos no tiene implicación clínica ya que hasta en el 90% de las mujeres esta infección es asintomática y transitoria que se resuelve durante los dos años posteriores a la infección (Castellsagué, 2008), gracias al desarrollo de una respuesta inmunitaria eficaz (Castellsagué et al., 2014). En el 10% de las mujeres la infección persiste y de estas, el 1% desarrollará un cáncer asociado al HPV (Alemany L. et al., 2016). En el caso de los hombres, la incidencia de la infección es similar a las mujeres pero la duración y persistencia de la misma es más corta (Giuliano et al., 2010).

Sin embargo, según Alemany y cols. el virus puede permanecer latente e indetectable para reactivarse posteriormente. Tanto hombres como mujeres pueden actuar como portadores y vehículos infectivos del virus durante largo tiempo (Alemany L. et al., 2016).

Existen unos factores de riesgo para el desarrollo de una infección por HPV persistente, la progresión de las lesiones de bajo grado a alto grado y el desarrollo de cáncer de cérvix, que son (Castellsagué, 2008), (Castellsagué et al., 2002), (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013): inicio precoz de las relaciones sexuales, promiscuidad sexual, nivel socioeconómico bajo, genotipo viral (siendo los genotipos 16 y 18 los más agresivos), la carga viral, la inmunodepresión (destacando la secundaria a la infección por el VIH), el hábito tabáquico, el uso

prolongado de anticonceptivos orales, elevado número de partos y la coinfección con otras infecciones de transmisión sexual como el herpes genital tipo 2 o la *Chlamydia trachomatis*.

España es considerado un país de bajo riesgo de desarrollo de cáncer de cérvix, con una prevalencia actual de la infección por HPV en mujeres es del 14%, variable entre comunidades autónomas, siendo superior al 15% en Navarra, Baleares, la Rioja y Murcia, y más baja del 10% en Castilla La Mancha, Castilla y León, Asturias y Cantabria (Álvarez García et al., 2024). Respecto a la distribución de la infección por rangos de edad, presenta la prevalencia más alta en jóvenes, con una prevalencia del 25% a los 25 años y una prevalencia del 5-10% a partir de los 30 años.

1.3.3. ETIOPATOGENIA

El HPV se introduce en el cérvix accediendo a las células basales del epitelio de la zona de transición escamo-columnar a través de microtraumas que alteran el epitelio de la mucosa, e integra su ADN con el de las células del epitelio (Sendagorta-Cudós et al., 2019). De esta forma altera la capacidad madurativa y adoptan alteraciones en su citología y arquitectura que se reflejan como atipias celulares que reflejan la displasia. El intervalo entre la exposición al virus y el desarrollo de la displasia oscila entre 6-24 semanas; aunque en la mayoría de las pacientes estas infecciones son autolimitadas eliminándose a los 8 meses de media, aunque hay algunas que se prolongan a más de dos años.

Es un virus que puede actuar sobre la célula huésped epitelial de 3 formas diferentes (**Figura 1.20**):

- Forma episomal. El virus se dispone su ADN viral de forma libre en el núcleo de las célula huésped, sin acoplarse al ADN huésped ni replicarse. No produce patología ni alteraciones celulares.

- Forma productiva no integrada. El virus se replica de manera independiente al ADN cromosomal de la célula huésped al que no se integra. El virus no sintetiza su cápside proteica hasta alcanzar los estratos superficiales del epitelio, que es donde manifiesta las características citopáticas del la infección como la atipia coilocítica, propia de las lesiones epiteliales escamosas de bajo grado (CIN 1) o los condilomas.

- Forma productiva-integrada. El virus integra su ADN en el de la célula huésped, sirviéndose del genoma de la célula infectada para replicarse donde produce alteraciones citológicas propias de las lesiones intraepiteliales escamosas de alto grado (CIN 2-3).

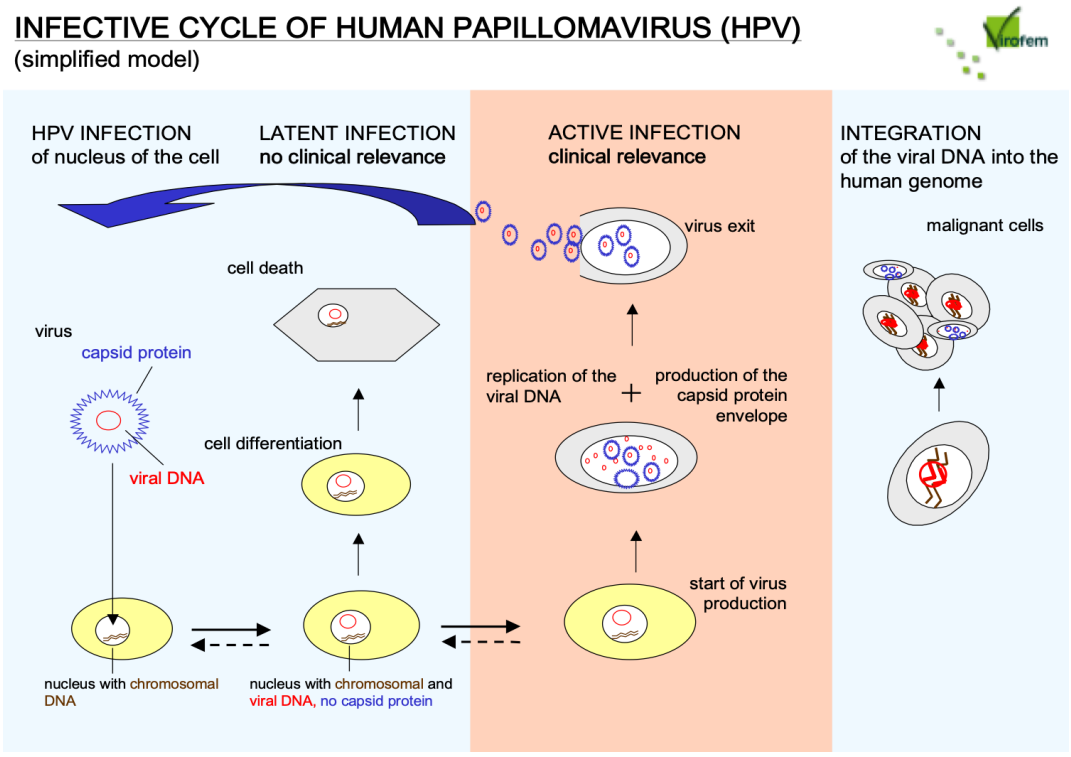


Figura 1.20. Ciclo infectivo del HPV. Imagen extraída de la comunicación "Aplicación de técnicas complementarias a la citología, para la detección del virus del papiloma humano (HPV) en lesiones epiteliales del aparato genital femenino", de García Macías.y cols. del V Congreso Virtual Hispanoamericano de Anatomía Patológica.

1.4. LA CITOLOGÍA CERVICOVAGINAL

1.4.1. HISTORIA Y FUNDAMENTOS BÁSICOS DE LA CITOLOGÍA CERVICOVAGINAL

En agosto de 1941 un médico griego afincado en Estados Unidos, llamado Dr. George Nicholas Papanicolau publicó junto al Dr. Traut en Nueva York en la revista *American Journal of Obstetrics and Gynecology* una comunicación sobre el valor del frotis vaginal en el diagnóstico del carcinoma de útero, titulada *"The diagnostic value of vaginal smears in carcinoma of the uterus"* (Papanicolaou & Traut, 1997). Recalca en su publicación que; a pesar de que un diagnóstico precoz da lugar a una mejora en el pronóstico, encontraban una gran dificultad para llegar a un diagnóstico precoz, ya que dependían únicamente de los subjetivos síntomas de enfermedad que percibía la paciente que le hacían consultar al médico. En muchas ocasiones, cuando la paciente notaba sintomatología para consultar a un especialista, la enfermedad estaba avanzada. De esta manera Papanicolau describió un sistema de recogida de muestra, establece una técnica de tinción de celular y asentó unos criterios de interpretación de resultados.

En su artículo describe una técnica barata y sencilla para diagnosticar el cáncer de cérvix en estadios tempranos: la citología cervicovaginal. Tras la observación de frotis vaginales normales y anormales de mujeres y animales durante muchos años, concluyó que las células del cérvix se disocian del epitelio y caen a la vagina, pudiendo ser recogidas y depositadas en cristales para ser estudiadas. Mientras estudiaba la maduración hormonal de la mucosa vaginal, descubrió que también se podrían identificar células tumorales en el flujo vaginal de las mujeres con cáncer de cérvix.

Las primeras citologías se tomaban con pipetas de cristal que aspiraban el fluido vaginal y lo depositaban posteriormente en 3 cristales diferentes, separando las células de vagina, exocérvix y endocérvix, conocida como la triple toma de Wied (Papanicolaou & Traut, 1997) (**Figura 1.21-A**). Antes de darle color a las células, descubrieron que había que someter la muestra a un proceso de fijación para

conservarlas, consistente en sumergir inmediatamente la muestra en una solución de alcohol al 95%.

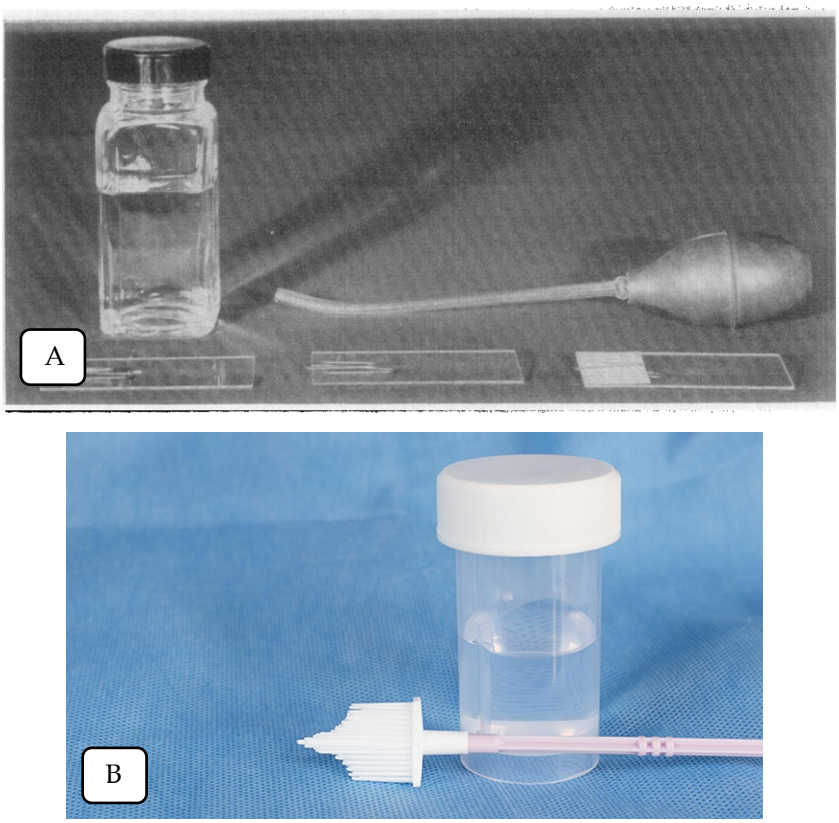


Figura 1.21. Comparativa de distintos métodos de elaboración de la citología. (A) Pipeta y botella que se usaban para la elaboración de la citología cervicovaginal descrita por Papanicolau, extraída de la publicación "The diagnostic value of vaginal smears in carcinoma of the uterus" (Papanicolaou & Traut, 1997). (B) Método actual de toma de citología líquida que emplea un cepillo cervical y líquido conservante celular, extraída de la web "Artemedica" (disponible en: https://artemedica.es/ginecologia/citologia-liquida-un-mejor-diagnostico-en-cancer-de-cuello-de-utero/).

Describen un método de tinción que empleaba 3 colorantes con distintas afinidades por radicales proteicos celulares, que se conoce como *tinción de Papanicolau*. Esta tinción tricrómica ha perdurado durante décadas y; excepto por discretas modificaciones, se mantiene hasta nuestros días. Emplea la *hematoxilina* de color azul que presenta afinidad por los radicales básicos presentes en el núcleo

celular y permite reconocer los cambios malignos celulares con precisión (Papanicolaou & Traut, 1997); la *EA* que tiene un color azul verdoso y el *Orange* de color naranja, ambos con afinidad por radicales ácidos del citoplasma.

Los resultados de las citologías los emitían en una clasificación de 5 categorías basadas en criterios diagnósticos y morfológicos, pero con algunas limitaciones técnicas con un alto porcentaje de falsos positivos y negativos. Además, dicha clasificación no permitía diferenciar las infecciones de alto riesgo y bajo. Dichas categorías son las siguientes (**Tabla 1.2**): (1) *Papanicolau I*: citología normal; (2) *Papanicolau II*: citología con inflamación (atípica pero sin evidencias de malignidad); (3) *Papanicolau III*: citología sospechosa para células neoplásicas (sugiere malignidad pero no es concluyente); (4) *Papanicolau IV*: citología con alta sospecha de malignidad y (5) *Papanicolau V*: citología con células claramente malignas.

Tabla 1.2. Tabla de equivalencia del antiguo Sistema de clasificación diagnóstica de Papanicolau con otros sistemas posteriores como el de Richart o el Sistema Bethesda empleado en la actualidad. Elaboración propia.

Descript.	Normal	Atipia benigna	Displasia leve	Displasia moderada	Displasia grave	Ca in situ	Ca invasor
Papanicolaou	I	II	III			IV	V
Richart	0		I	II	III		Ca. invasor
Bethesda	Normal	ASCUS	LSIL	HSIL			Ca. invasor

Tras su publicación, este sistema no fue aceptado por la comunidad científica porque no se conocía con precisión los cambios celulares asociados al cáncer ni la historia natural del mismo. Con los años, gracias al conocimiento de la etiopatogenia del cáncer, este sistema y su tinción se han establecido como la mejor técnica de cribado poblacional disponible en la actualidad. Papanicolau sentó las bases de la toma de muestra, tinción y clasificación que con el tiempo han dado paso al Sistema Bethesda, aceptado en numerosos laboratorios del mundo.

1.4.2. ELABORACIÓN DE LA CITOLOGÍA CERVICOVAGINAL

Este tipo de estudio descrito por Papanicolau ha pasado a conocerse como citología exfoliativa, que hace referencia a que las células obtenidas se han desprendido de la superficie epitelial de manera espontánea. Para ayudar a la recogida de estas células el profesional se ayuda de un cepillo que recoge células del fondo vaginal, exocérvix y endocérvix (**Figura 1.21-B**), para asegurarse de que se obtienen células de la UEC por ser la más vulnerable a la infección por el HPV y; por lo tanto, al desarrollo de lesiones preneoplásicas y neoplásicas. Tras la toma de la muestra, su procesamiento y tinción, el patólogo estudia estas citologías al microscopio, y analizando características citológicas, es capaz de determinar el estado del tejido, permitiendo detectar patología como lesiones inflamatorias, infecciosas, displásicas o tumorales.

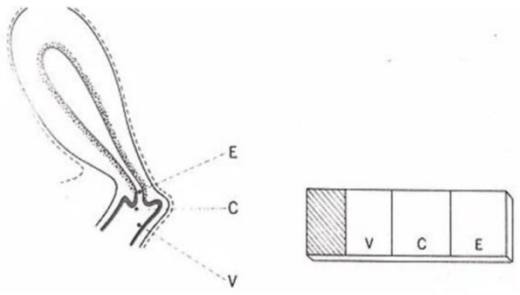


Figura 1.22. Método de recogida de muestra mediante la triple toma o citología convencional. De forma separada se toma la muestra del endocérvix (e), exocérvix (c) y vagina (v) y se depositan de manera independiente en 3 territorios del portaobjetos bien delimitados. Imagen obtenida de https://citomol.com/citologias/.

En la actualidad hay dos sistemas de elaboración de la citología cervicovaginal: la citología convencional (o triple toma) y la citología en medio líquido.

En la citología convencional se obtiene de manera independiente la muestra de fondo de saco vaginal, exocérvix y endocervix para extenderla directamente en un portaobjetos en 3 zonas separadas y diferenciadas para cada territorio

muestreado (**Figura 1.22**). La calidad de este tipo de citología viene muy condicionada por la destreza del profesional, ya que si no se realiza una adecuada toma y extensión de la muestra, se puede empeorar su calidad (con grumos y artefactos) o su representatividad por no tener células de la UEC.

En la citología líquida la muestra se obtiene mediante un cepillo especial de cerdas flexible y mango desmontable denominado *Liquid-Based Cytology Specimen brush,* que recoge las células de las distintas zonas a estudio, para posteriormente pasarlas a un medio líquido conservante celular de base alcohólica que actúa como fijador (MultiCyte® ThinPrep®) y permite su almacenaje y transporte (**Figura 1.21-B y 1.23**) (Ronco et al., 2007), (Arbyn et al., 2008). Tras su procesamiento y centrifugación, el medio líquido se filtra para depositar las células sobre un área limitada del portaobjetos en la que se disponen las células en una monocapa, proceso que generalmente es automatizado (PrepStain® o SuperPap®) (**Figura 1.24**). Tras su tinción con la técnica de Papanicolau en la que las células adquieren unos colorantes que permiten su visionado, las células pueden ser diagnosticadas por un patólogo. Esta mejoría en la calidad del muestra disminuye el número de citologías no adecuadas para diagnóstico (Torné A. et al., 2022). Dentro de los sistemas de citología líquida disponibles, dos han sido validados por la FDA y por distintos ensayos clínicos: el sistema ThinPrep® de *Hologic* y el sistema SurePath® de *BD* **(Figura 1.25**).

Ambos métodos están aceptados para el cribado del cáncer de cérvix, aunque la citología líquida es preferente en la mayoría de guías clínicas por tener numerosas ventajas sobre la citología convencional. Entre las ventajas más destacables encontramos que la citología líquida es más limpia, ya que elimina gran parte de la celularidad de fondo que entorpece su interpretación, como hematíes y polimorfonucleares neutrófilos. Además, al pasarse la muestra a un medio líquido, hace posible su automatización en el procesamiento del laboratorio, su lectura automatizada por sistemas de inteligencia artificial y su aprovechamiento con la utilización del mismo material para la determinación del genotipado HPV (Torné A. et al., 2022), (Lee et al., 1997).

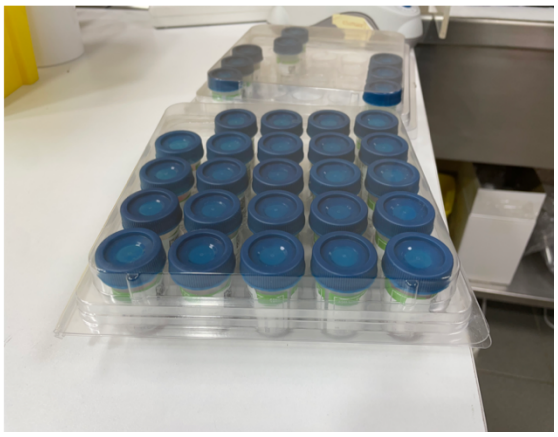


Figura 1.24. Líquido conservante celular que se emplea para la fijación de la muestra en la citología liquida junto a su cepillo especial para la recogida de la muestra. Elaboración propia.

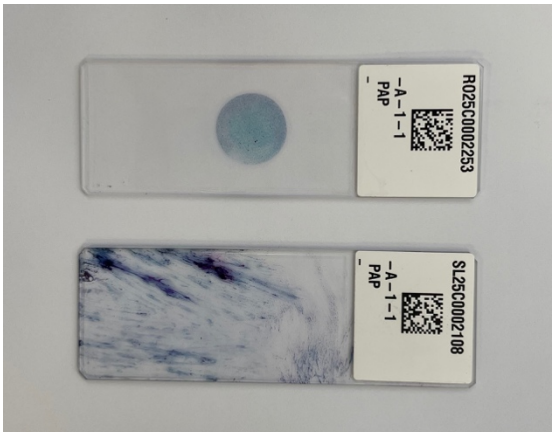


Figura 1.23. Comparativa del portaobjetos con muestra obtenida mediante citología líquida (arriba) y triple toma (abajo). Nótese que en la citología líquida todas las células de todos los territorios muestreados se disponen juntos en un área redondeada bien delimitada. Por el contrario, en la triple toma, se observan distintas secciones lineales en el cristal, cada una correspondiente a cada uno de los territorios muestreados de manera independiente. Elaboración propia.

Existen numerosos estudios que demuestran que la sensibilidad de la citología convencional es baja (de aproximadamente el 50%) ya que se ve influenciada por factores como la calidad en la toma de la muestra, su conservación y extensión. Por el contrario, otros estudios reflejan que la citología líquida no solo ha aumentado ligeramente la sensibilidad en los laboratorios, sino que acorta los

tiempos de diagnóstico por tratarse de una determinación de mejor calidad (de Bekker-Grob et al., 2012) (Hutchinson et al., 1999). Una revisión sistemática de la *Cochrane library* reveló una sensibilidad en el diagnóstico de HSIL / CIN2 del 89,9% en la prueba HPV, del 62,5% en la citología convencional y del 72,9% en la citología líquida, y una especificidad del 89,9%, 96,6% y 90,3%, respectivamente (Koliopoulos et al., 2017).

Para la tinción de las citologías se emplea la tinción que describió Papanicolau con discretas modificaciones (**Tabla 1.3**). Actualmente está muy extendido el empleo de sistemas automatizados de procesamiento y tinción de las citologías (**Figura 1.25**), que consta de los siguientes pasos: se inicia introduciendo la muestra en concentraciones decrecientes de alcohol, pasándola por una solución de hematoxilina de Harris y ácido clorhídrico al 0,5%. Posteriormente se introduce por una batería de soluciones ascendentes de alcohol y se tiñe con la solución de Papanicolau OG y EA, para finalmente deshidratar la muestra con etanol al 96% y absoluto, aclarar y montar tras pasar por distintos lavados para eliminar el exceso de colorante.

Figura 1.25. Maquinara SuperPap® empleada en la elaboración automatizada de la citología líquida. Elaboración propia.

1.4.3. DIAGNÓSTICO DE LA CITOLOGÍA CERVICOVAGINAL CON EL SISTEMA BETHESDA

Con los avances en los conocimientos de la carcinogénesis del cáncer de cérvix, se planteó la necesidad de unificar terminologías para facilitar la comunicación entre médicos y proporcionar un tratamiento optimizado a las pacientes impulsados por hallazgos observados en su citología. Es así cuando entre 1988-1991 se reunió un grupo de expertos en el Instituto Nacional de Salud de Bethesda en Maryland en Estados Unidos, para asentar las bases del *Sistema Bethesda de diagnóstico de la citología cervicovaginal* (Pangarkar, 2022). En esta reunión (1) se evaluó el impacto de la clasificación, (2) se propusieron algunas reformas, (3) se decidió no emplear más la clasificación de Papanicolau con grados numéricos y (4) se propuso emplear términos diagnósticos claros que no generaran ambigüedades entre el patólogo y el ginecólogo. Posteriormente en 1991, 2001 y 2014 se publicaron algunas actualizaciones y hoy día está implantado en laboratorios de todo el mundo (Nayar & Wilbur, 2015).

Tabla 1.3. Componentes de la tinción de la Papanicolau y funcionalidad de cada uno de los colorantes que la integran. Elaboración propia.

Tinción	Componentes que contiene	Funcionalidad
Solución de hematoxilina	Hematoxilina de Harris	Tiene afinidad por los núcleos empleando el pigmento Hematoxilina de Harris.
Solución de Papanicolau OG	Orange G	Colorante ácido sintético con afinidad por los compuestos básicos, tiñendo la célula de color naranja brillante.
Solución de Papanicolau EA	Eosina Y	Tiñe los citoplasmas de color rosa-anaranjado
	Luz verde SF	Las células escamosas de estratos bajos e intermedios, se tiñen de color verde azulado, por lo que marca las células inmaduras o en maduración.
	Pardo Bismark	Es un colorante que tiñe las mucinas citoplasmáticas resaltando las células epiteliales productoras de mucina como las endocervicales y endometriales.
	Ácido fosfotúngstico	Es un mordiente que aumenta la afinidad de las células por el colorante aumentando la intensidad de coloración de las células escamosas inmaduras.

El Sistema Bethesda analiza todos los elementos que se pueden encontrar en la citología de Papanicolau y establece una categorización general e interpretación en el diagnóstico de citología cervicovaginal. En el Sistema se define la "*lesión intraepitelial escamosa o SIL*", definiendo los cambios celulares que experimente la célula escamosa displásica tras su infección por el HPV. Dentro de esta categorización general, encontramos los siguientes conceptos (Pangarkar, 2022).:

- *Citología normal*: células benignas sin alteraciones citológicas significativas (**Figura 1.26**).

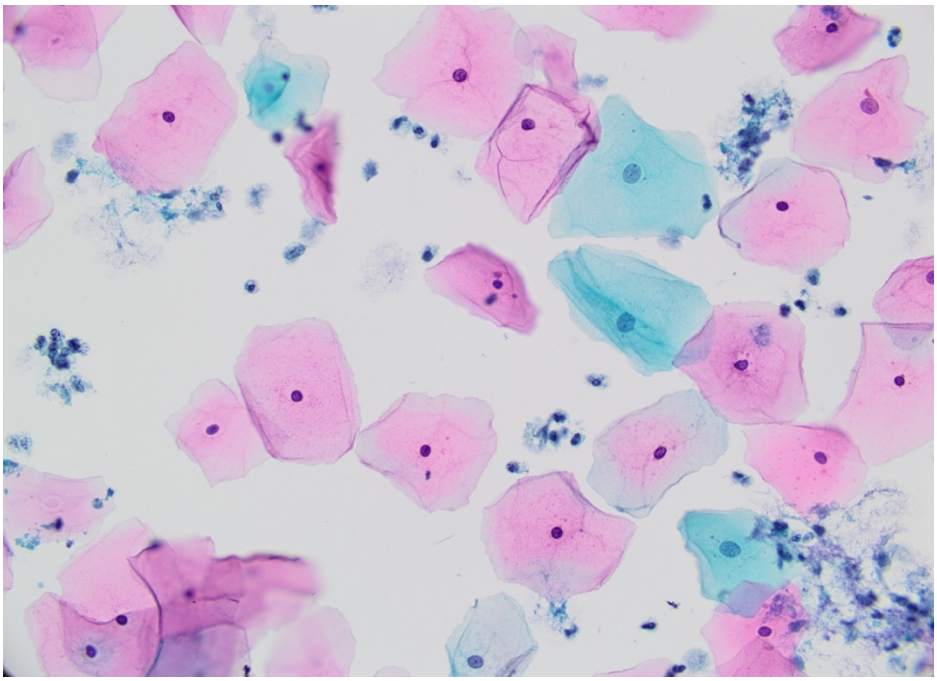


Figura 1.26. Células escamosa cervicales benignas. Nótese una baja relación de la relación núcleo citoplasma. Podemos distinguir células exocervicales de estratos superficiales (de coloración orangófila o rosada) e intermedios (de coloración cianófila o azulada). Elaboración propia.

- *Atipia de células escamosas de significado incierto o ASC-US*. Se observan alteraciones leves sospechosas pero que no son suficientes para catalogarlo como lesión intraepitelial de bajo grado. Incluyen aquellas alteraciones celulares que son mayores a las esperadas por cambios reactivos pero insuficientes para catalogarlas como lesión intraepitelial (Solomon et al., 2002), (Nayar & Wilbur, 2015). La presencia de núcleos con un tamaño 2-3 veces el tamaño de una célula intermedia normal, un aumento leve de la

relación núcleo-citoplasma, una hipercromasia mínima o una ligera irregularidad en la membrana nuclear, permiten catalogar la citología como ASC-US (**Figura 1.27**).

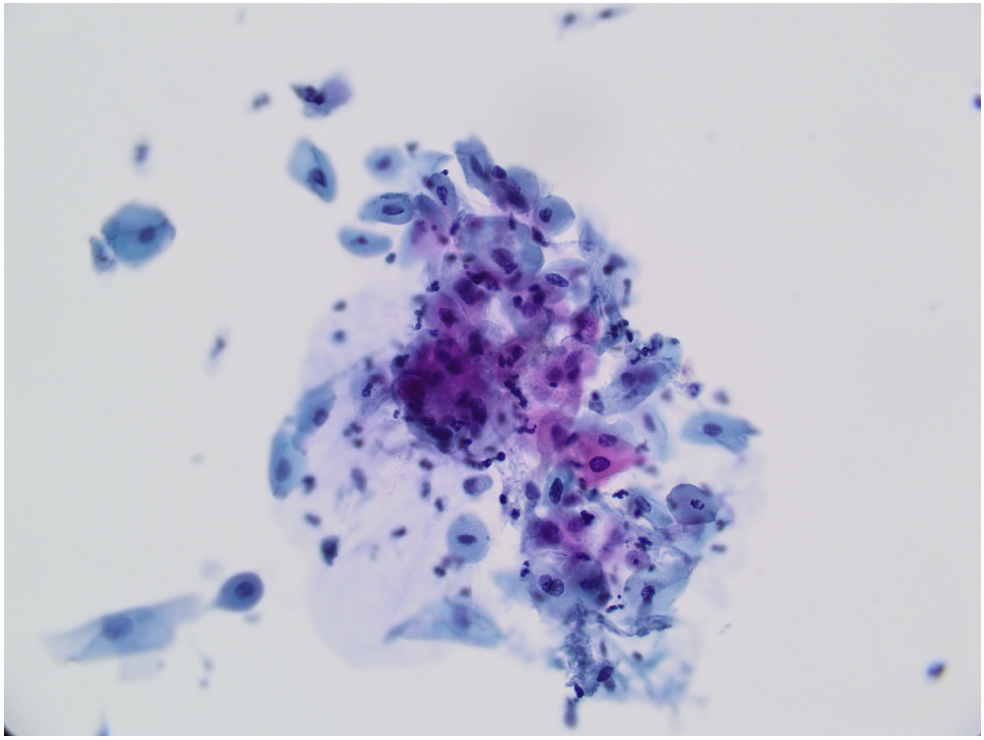


Figura 1.27. Células con ligero aumento de la relación núcleo-citoplasma e irregularidad nuclear ligera, que no son suficientes para catalogarlo de una lesión de mayor agresividad. Elaboración propia.

- *Lesión escamosa intraepitelial de bajo grado o LSIL*. En LSIL se observan cambios en el tamaño y forma nuclear que sugieren que hay una infección por HPV que no se ha integrado en el ADN del huésped. Este tipo de lesiones afectan a las células escamosas de estratos altos del epitelio, conocidas como células superficiales o intermedias, y se corresponden con displasias de leve intensidad (CIN 1) en los estudios por biopsia. Los cambios citológicos que sugieren LSIL son de manera resumida (Pangarkar, 2022): 1) un aumento del tamaño nuclear de al menos 3 veces el tamaño de una célula intermedia normal con el consecuente aumento leve de la relación núcleo-citoplasma; 2) variedad en el tamaño nuclear y 3) binucleación, multinucleación y presencia de halos blanquecinos

perinucleares, conocidos como coilocitos **(Figuras 1.28 y 1.29)**. Las lesiones LSIL tienden a resolverse espontáneamente gracias a la acción del sistema inmunitario de la paciente.

- *Lesión intraepitelial escamosa de alto grado o HSIL*. Se observan cambios nucleares que indican que el virus HPV ha integrado su ADN en el de la célula huésped. Los cambios nucleares que sugieren HSIL son de manera resumida (Pangarkar, 2022): 1) un aumento del tamaño del núcleo de una célula pequeña (de estratos bajos) que se corresponde con un marcado aumento de la relación núcleo-citoplasma; 2) contornos nucleares irregulares con frecuentes indentaciones y 3) hipercromatismo nuclear (aunque puede ser normo o hipocromático) (**Figura 1.30 y 1.31**). En este caso las alteraciones nucleares afectan a todos los estratos del epitelio, que se corresponde con una displasia moderada o intensa (CIN 2 y 3) en el estudio por biopsia.

- *Células escamosas atípicas, no se puede descartar HSIL (ASC-H*). Se observan alteraciones sospechosas pero no diagnósticas de lesión intraepitelial escamosa de alto grado.

Las categorías diagnósticas intermedias ASC-US y ASC-H se reservan para los casos inciertos con hallazgos sospechosos pero no certeros de displasia; y las categorías LSIL y HSIL cuando se observan hallazgos certeros de displasia.

Este sistema de clasificación permite predecir el comportamiento de la displasia y por lo tanto planificar un seguimiento de la paciente. Según Torné y cols. (Torné A. et al., 2022) entre el 60-80% de las lesiones LSIL (CIN 1) se resuelven solas sin tratamiento y un 5-10% evolucionan a un HSIL, independientemente del genotipo del HPV. Según McCredie y cols,. las lesiones HSIL (CIN3) progresan a cáncer en el 31,3% en 30 años (McCredie et al., 2008). En el caso de las lesiones CIN1, se ha descrito que el 70% regresan en 6 años y un 6% progresa a HSIL (CIN3) o cáncer de cérvix (Holowaty et al., 1999), (Melnikow et al., 1998).

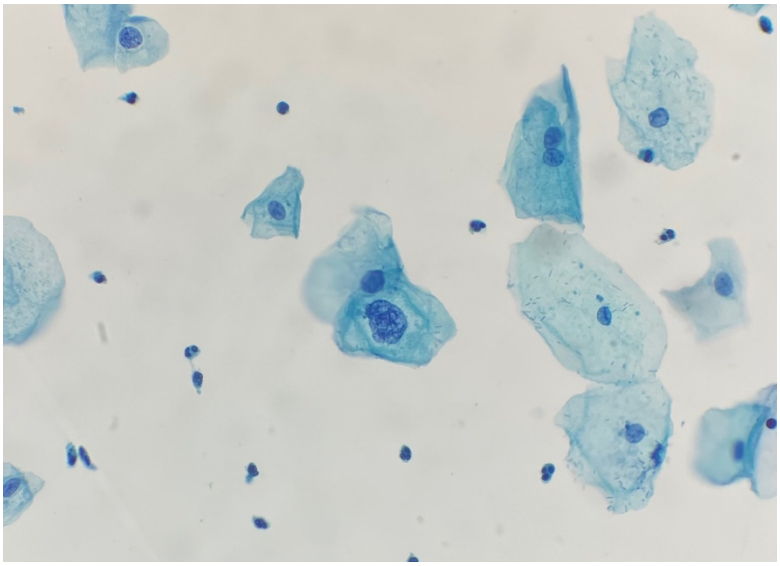


Figura 1.29. Se observa una célula epitelial de tipo intermedia con alteraciones citológicas marcadas con un aumento de la relación núcleo-citoplasma e irregularidad nuclear. Elaboración propia.

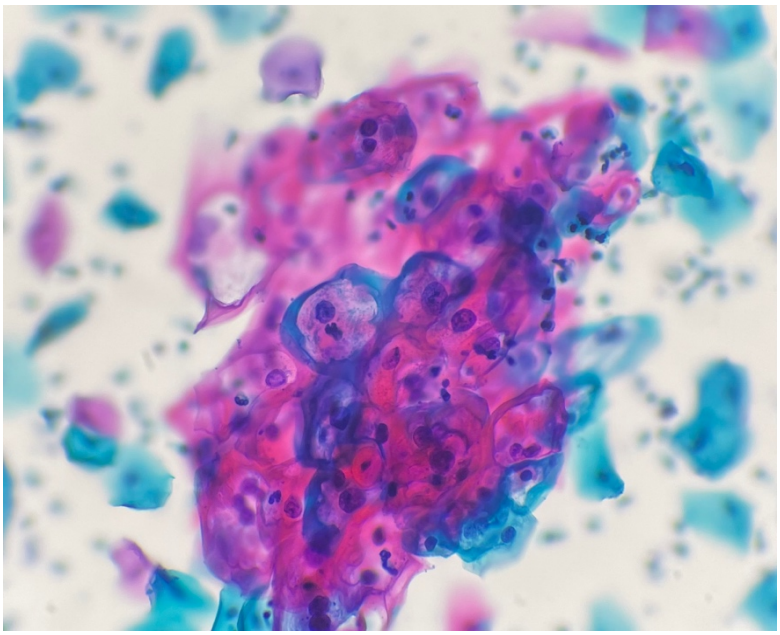


Figura 1.28. Placa de coilocitos LSIL en la que se observa binucleación, aumento de la relación núcleo-citoplasma con irregularidad e hipercromasia nuclear, con un halo blanquecino perinuclear amplio, que define a los coilocitos. Elaboración propia.

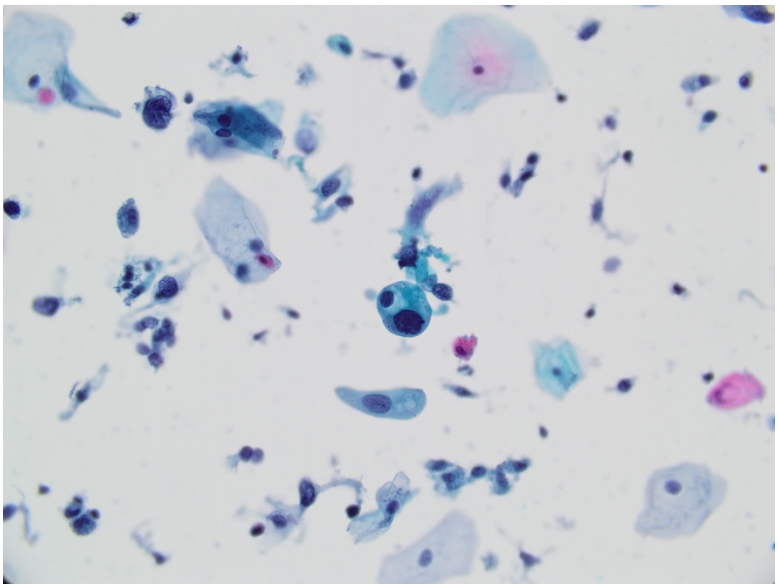


Figura 1. 30. Se observa en el centro dos células epiteliales del estrato basal o parabasal con un aumento marcado del tamaño nuclear, con intenso hipercromatismo e irregularidad nuclear. Elaboración propia.

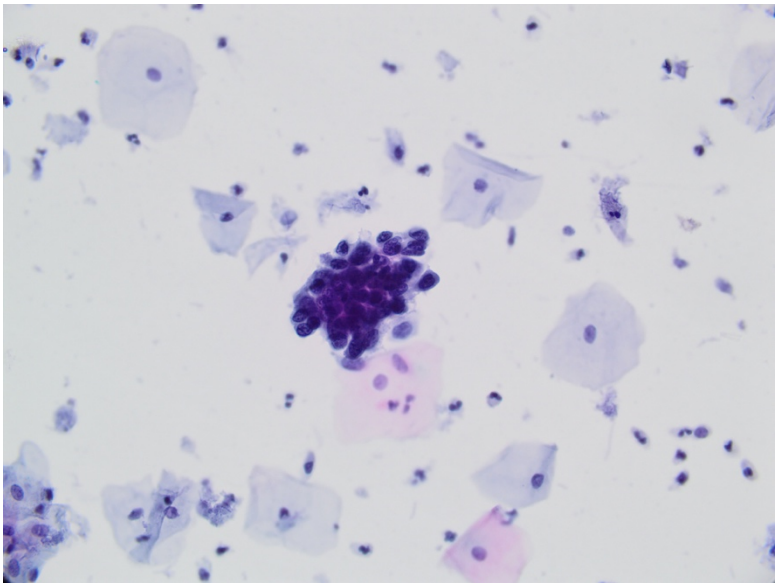


Figura 1.31. Placa de células con HSIL. Elaboración propia.

1.4.4. LIMITACIONES Y FRONTERAS EN EL DIAGNÓSTICO CITOLÓGICO

La citología cervicovaginal lleva un control de calidad dependiente del procesamiento, tinción, interpretación y notificación del diagnóstico; para la obtención de una acreditación de calidad (norma *UNE-EN ISO 15189)* (Torné A. et al., 2022). Debe emplear un personal cualificado que elabore un número de citologías suficiente como para mantener una buena experiencia (Torné A. et al., 2022). Medidas como el desarrollo de la citología líquida, la automatización en el laboratorio y el seguimiento de guías diagnósticas, tratan de mejorar la calidad del procedimiento.

A pesar de que los criterios de diagnóstico están unificados en el Sistema Bethesda, el diagnóstico citológico está sujeto a una variación inter e intrapersonal debido a que presenta un cierto componente subjetivo en el que influyen la fatiga del observador, la calidad de la muestra y los artefactos de laboratorio, entre otros (Bigras et al., 2013) (Ismail et al., 1989) (Stoler et al., 2001). También se trata de una prueba con alto porcentaje de falsos negativos (hasta el 40%) dependientes de la calidad en la toma de la muestra.

Según distintos autores la sensibilidad de la técnica varía entre 68,8 % y el 93,8 %; la especificidad entre el 70,6 % y el 95,6 %; y presenta una fiabilidad en la detección de HSIL (CIN 2 o mayores) que oscila entre el 74,1 % y el 83,8 % (Bigras et al., 2013) (Arbyn et al., 2004) (Sørbye et al., 2017) (Wright et al., 2014). La subjetividad en la interpretación y los posibles errores técnicos humanos justifican esta variabilidad de los resultados consultados.

Uno de los avances más importantes en los últimos años ha sido el desarrollo de la citología líquida y de la lectura automatizada (Torné A. et al., 2022). La citología líquida permite obtener muestras de mayor calidad, con mejor visualización y menos artefactos, que facilitan el diagnóstico (Torné A. et al., 2022). La lectura automatizada se presenta como una vía prometedora de ayuda al diagnóstico citológico, en la que se basa este proyecto.

1.5. ESTADO DEL ARTE DE LA INTELIGENCIA ARTIFICIAL

1.5.1. HISTORIA Y CRONOLOGÍA EN LOS SISTEMAS DE LECTURA AUTOMATIZADA DE CITOLOGÍAS CERVICOVAGINALES

SISTEMAS DE DETECCIÓN TEMPRANA

El primer sistema automatizado de análisis de citologías teñidas con la tinción de Papanicolau registrado fue el *Cytoanalyzer* (Tolles, 1955), que distinguía las células cancerosas de las normales a través del análisis de la densidad óptica y tamaño nuclear de las células. Mediante este punto de partida, los sistemas de primera generación que se desarrollaron posteriormente se basaron en este concepto básico de que las células cancerosas adquirían unas características en el núcleo que permitían diferenciarse de las células normales. Por ejemplo, los sistemas *TICAS* (Wied et al., 1975) y *CYBEST* (Watanabe, 1974) generaban un histograma que mostraba las diferencias en la morfometría entre las células cancerosas y las normales, mediante circuitos de procesamiento de vídeos analógicos cableados. Sin embargo, estos sistemas no disponían de ordenadores que permitieran la visualización de las imágenes digitales disponibles y se generaban resultados falsos.

En la década de 1980 se dio paso a los sistemas de detección de segunda generación como los sistemas *Bio-PEPR*, *FAZYTAN*, *Cerviscan*, *LEYTAS* y *Discanner* (Bengtsson & Malm, 2014). La problemática a la que se enfrentaron los sistemas de segunda generación era que los ordenadores eran lentos y no podían procesar toda la citología de Papanicolau, que podría contener hasta 300.000 células. También encontraban dificultades para detectar los límites celulares al trabajar con grupos tridimensionales de células. Estos sistemas marcaron un comienzo que permitió explorar la segmentación de imágenes, extracción de características y la clasificación de la citología.

SISTEMAS COMERCIALES DE PRIMERA GENERACIÓN

A finales de 1980 con el aumento de la demanda de citologías cervicovaginales se observó una gran disminución en la calidad del diagnóstico a expensas de examinar un gran número de portaobjetos. Se establecieron directrices importantes como las Enmiendas de Mejora de Laboratorios Clínicos en 1988 y la primera edición de los Criterios del Sistema Bethesda (Nayar & Wilbur, 2017). También se produjo un gran avance en el sistema informático de los ordenadores aumentando la velocidad de procesamiento, memoria y visualización de imágenes, aumentando las esperanzas de automatización en el cribado de citología cervicovaginal.

En la década de los 90 la FDA aprobó tres sistemas para la lectura automatizada de citologías (Neopath, EE. UU.), (Neuromedical Systems, EE. UU.) (Koss et al., 1994), (Patten et al., 1997), (Lew et al., 2021), (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013). Ninguno de estos sistemas iniciales llegó a comercializarse.

- Para la citología convencional: *AutoPap 300QC* y *PapNet.* El sistema *AutoPap 300QC* (Patten et al., 1996) era un procesador de imágenes que daba una puntuación total acumulativa del portaobjetos, que cuando alcanzaba una puntuación alta era revisadas por un citotécnico (**Figura 1.32**). *PapNet* (Koss et al., 1994) era una red neuronal que seleccionaba hasta 128 imágenes de células sospechosas para que el citotecnólogo las visualizara posteriormente en un monitor (**Figura 1.33**).

- Para la citología líquida *AutoCyte Screen*.

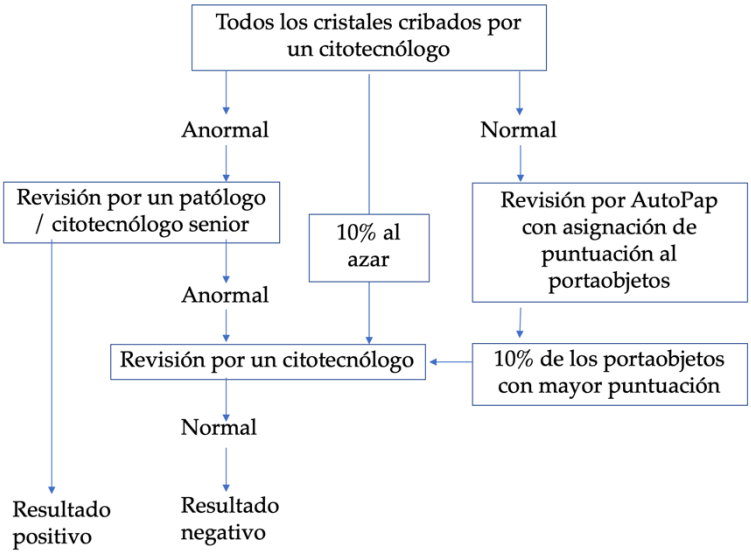


Figura 1.32. Flujo de trabajo que muestra la inserción de AutoPap 300 QC en el flujo de trabajo del laboratorio del hospital. Elaboración propia.

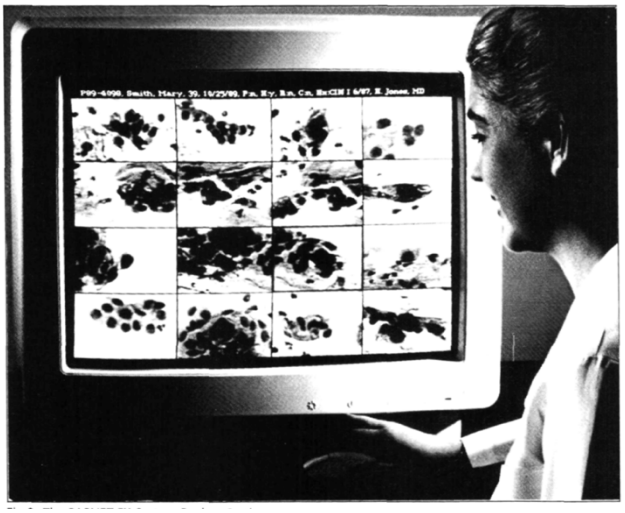


Figura 1.33. Monitor de visionado de las células potencialmente anormales del sistema PapNet. Imagen obtenida de la revisión sistemática: "A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis" de Jiang y cols.(Jiang et al., 2023).

Con el desarrollo en 1997 del sistema *PapNet*, se asentó la base de posteriores sistemas automatizados que sí se comercializaron y que se han ido desarrollando con los años, como *Hologic ThinPrep ® Imaging System (Imager)* o *BD FocalPoint ™ Slide Profiler (*que sustituye al sistema de primera generación *AutoPap 300 QC)* (Koss et al., 1997) , (Bengtsson & Malm, 2014), (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013).

1.5.2. COMIENZOS DE LA INTELIGENCIA ARTIFICIAL

En las últimas décadas del siglo XX y primeras del XXI, la tecnología ha ido desarrollándose y perfeccionándose para responder a la necesidad de gestionar grandes cantidades de información, dando así al desarrollo de la *inteligencia artificial (IA).* La IA es un universo muy amplio que engloba muchas áreas con las que se superpone, como la minería de datos, neurocomputación, reconocimiento de patrones o el machine learning (del inglés, aprendizaje automático) (Miller & Brown, 2018). El término IA se emplea como un concepto general que recoge todos los métodos que emplea un ordenador para simular o superar a la inteligencia humana (Hamet & Tremblay, 2017).

El objetivo de la IA, de forma general, se basa en que las máquinas sean capaces de aprender y actuar de forma similar a como lo haría un ser humano, con capacidad de aprendizaje autónomo mediante la incorporación de datos nuevos e información del mundo real, manejada por seres humanos. Es decir, trata de crear sistemas informáticos que son capaces de interpretar la información como si se tratara de un ser humano, dando predicciones y trabajando de forma semiautónoma (Chang et al., 2019).

Desde los primeros avances de la IA se ha tratado avanzar en el ámbito médico. En 1966 se propuso un chatbot llamado *ELIZA* (Weizenbaum, 1966) que podía interactuar con pacientes como un psicoterapeuta. En 1977 un sistema experto biomédico llamado *MYCIN* podía analizar los síntomas infecciosos para sugerir bacterias causantes y recomendar un tratamiento farmacológico (Shortliffe,

1977). En 1922 se desarrolló el *PATHFINDER*, un sistema equipado con razonamiento probabilístico en el diagnóstico en hematopatología (Heckerman et al., 2018), (Heckerman & Nathwani, 1992).

Aplicado al campo de la medicina se han potenciado técnicas como el *aprendizaje automático o machine learning (ML),* en inglés. El aprendizaje automático es un área de la IA que emplea métodos estadísticos para optimizar modelos para llevar a cabo una tarea específica sin depender de instrucciones humanas específicas que definan todas las normas o parámetros del modelo (Acs et al., 2020). Es decir; son sistemas que pueden aprender automáticamente e ir mejorando con la experiencia, sin necesidad de modificar el algoritmo durante el proceso (Rojo, 2019). En sus inicios estos sistemas se empleaban como logaritmos de estudio probabilístico para tratar de predecir la posibilidad de que un fenómeno ocurriera o no, en base a un entrenamiento previo concreto y a unos datos previos que permitían ponderar la relevancia de cada dato y su influencia en la ocurrencia del fenómeno a estudio.

Las técnicas de ML están muy extendidas en la tecnología de la sociedad actual, disponible en la mayoría de los dispositivos electrónicos, en el reconocimiento de voz, filtros de búsquedas en internet, reconocimientos de imágenes o sistemas de conducción automática, entre otros.

El *aprendizaje supervisado* es un tipo de ML que se basa en entrenar al modelo optimizando su rendimiento a través de la acción del ser humano empleando datos de entrenamiento; como por ejemplo, imágenes médicas con etiquetas disponibles como: datos clínicos, resultados de pacientes o anotaciones en las imágenes a nivel de píxeles. El algoritmo analiza cada muestra y le asigna una etiqueta o categoría según sus características tras un entrenamiento previo con una muestra similar, realizando ajustes internos de cálculo y ajustando parámetros. Cuando el algoritmo alcanza un número de ejemplos de muestra adecuado, será capaz de predecir con una probabilidad concreta, cual será la etiqueta más adecuada en las siguientes muestras que se le presenten. Estos algoritmos dominan este tipo de aplicaciones en el mundo de la medicina y el deep learning (Acs et al., 2020). De forma general,

estos sistemas se pueden agrupar en dos tipos de clasificaciones: binaria o multiclase. La clasificación binaria permite clasificar únicamente datos en dos categorías de etiquetado y la clasificación multiclase es capaz de discernir entre varias categorías distintas.

Un tipo de ML con una gran aceptación hoy día son las *redes neuronales artificiales*, que consisten en modelos matemáticos flexibles inspirados en el cerebro humano formado por una serie de neuronas, que trabajan con grandes volúmenes de datos y son capaces de identificar relaciones no lineales y de aprender de errores corregidos previamente, para mejorar la predicción del modelo (Miller & Brown, 2018). Aplicados a la medicina, estos métodos de reconocimiento de patrones que emplean redes neuronales artificiales de múltiples capas, permiten incorporar datos clínicos, radiológicos y genómicos para el diagnóstico y pronóstico de enfermedades. Estas redes neuronales artificiales constituidas en múltiples capas se conocen como *aprendizaje profundo.*

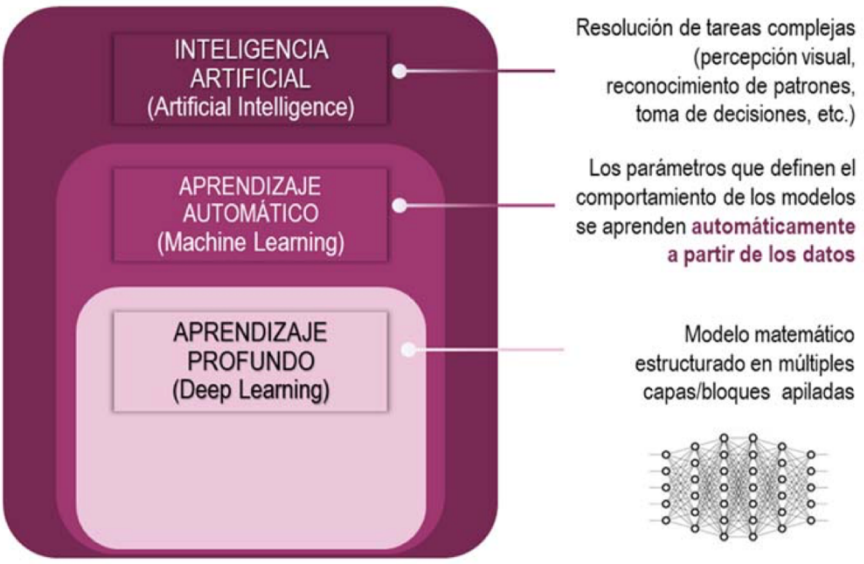


Figura 1.34. Relación entre la inteligencia artificial, el aprendizaje automático y el aprendizaje profundo. Imagen tomada de Marcial García Rojo Aplicaciones de deep learning, un método de inteligencia artificial, en Anatomía Patológica (Libro Blanco 2021).

El *aprendizaje profundo o deep learning (DL)* es un subcampo del ML basado en redes neuronales artificiales de varias capas de profundidad anidadas de neuronas (LeCun et al., 2015) **(Figura 1.34**). Estas redes son modelos que se inspiran en el procesamiento de información de redes neuronales biológicas que surgieron en 1940 (McCulloch & Pitts, 1943). Los modelos de DL tienen una capa de entrada (datos de imagen), varias capas ocultas y una capa de salida. A partir de 2006 Hinton y cols. diseñaron una red neuronal muy profunda que presentaba un aprendizaje no supervisado por el ser humano, que dio pie a que estas técnicas tuvieran un gran éxito y comenzaran a ser validadas en problemáticas reales del día a día como redes sociales o búsquedas en internet, y pasaron a tener un reconocimiento internacional (Miller & Brown, 2018). Estos algoritmos son capaces de solventar problemas muy complejos, que manejan con un gran volumen de datos con más de 5000 variables diferentes (Miller & Brown, 2018).

La historia del éxito del DL comienza en el campo de reconocimiento visual de objetos que a gran escala se desarrolló con ImageNet en 2012, que demostró un gran rendimiento de la red neuronal convolucional (Chang et al., 2019). El DL está presente en nuestros días en numerosas aplicaciones que ayudan a resolver problemas cotidianos según Le Cun y cols. (LeCun et al., 2015) y Jurtz y cols. (Jurtz et al., 2017). En el campo de la medicina, antes de la era del DL, se empleaban métodos de ML. Con la llegada de imágenes médicas digitales en Radiodiagnóstico como la resonancia magnética, la tomografía computarizada o los rayos X, se permitió el análisis de imágenes digitales donde la IA obtuvo grandes resultados. En 1994 Vyborny y Giger (Vyborny & Giger, 1994) desarrollaron algoritmos de ML en análisis de mamografías para la detección de masas tumorales, microcalcificaciones y la distinción de lesiones benignas y malignas. Demostraron la mayor eficacia de la detección asistida por computadora (CAD) frente al análisis de radiólogos únicamente. Los algoritmos de DL tienen un rendimiento diagnóstico semejante o mejor al de expertos humanos, como se ha demostrado en el análisis de imágenes de lesiones cutáneas en el diagnóstico de carcinomas de células escamosas frente a queratosis seborreicas benignas (Esteva et al., 2017). En otro estudio se obtuvieron grandes resultados en el diagnóstico de enfermedades retinianas a través de imágenes de tomografía de coherencia óptica (De Fauw et al.,

2018) o detección de retinopatía diabética a través de imágenes de fondo de ojo (Abràmoff et al., 2018).

Las *redes neuronales convolucionales (CNN*) son un tipo de modelos de DL que se usan para modelar datos sobre imágenes imágenes (Acs et al., 2020). Cuando el modelo se ha optimizado con gran cantidad de datos de entrenamiento, es capaz de aplicar los patrones aprendidos para proporcionar predicciones de respuestas o etiquetas en observaciones nunca vistas antes (Janowczyk & Madabhushi, 2016).

El acceso a grandes volúmenes de datos en red (cloud networking) ha hecho posible desarrollar este tipo de redes neuronales profundas de múltiples capas con diseños complejos, que son las más empleadas en el análisis de imágenes en Anatomía Patológica. Tras el éxito de la CNN profunda en la clasificación de imágenes, se ha tratado de aplicarlos a la medicina. En 2016 Gulshan y cols. (Gulshan et al., 2016) analizaron las imágenes de fondo de ojo retiniano mediante DL basado en CNN para detectar lesiones de retinopatía diabética, alcanzado una sensibilidad del 97,5 y especificidad del 93,4 %, y área bajo la curva de 0,991. Las CNN no son las únicas redes complejas, también encontramos las *redes neuronales recurrentes*, los *autocodificadores* o las *redes degenerativas adversarias* (Miller & Brown, 2018), (Robertson et al., 2018).

1.5.3. FUNCIONAMIENTO DEL DEEP LEARNING

Los sistemas de clasificación automática han ido mejorándose hasta desarrollar algoritmos cuyo aprendizaje es superponible al que realiza un ser humano, siendo capaces de clasificar imágenes directamente sin necesidad de un procesamiento previo, constituyendo una ventaja y diferencia fundamental con técnicas clásicas de ML. El ser humano aprende de manera intuitiva observando los fenómenos y extrayendo de ellos características comunes.

Por ejemplo, un niño aprende a diferenciar un perro de un gato, no por sus diferencias biológicas, sino por la imagen de ambos animales, entre las cuales el

niño trata de identificar las semejanzas y diferencias entre ellas. Su capacidad para saber diferenciarlos dependerá de la cantidad de perros y gatos de distintas razas, pelajes y colores que haya visto. El cerebro del niño durante el aprendizaje desarrolla sinapsis entre distintas neuronas y elimina otras, dependiendo de que la clasificación final sea correcta o no, seleccionando los circuitos sinápticos neuronales que obtengan mejores resultados. Esta capacidad de selección de circuitos neuronales es una característica cerebral conocida como plasticidad neuronal, capacidad que disminuye con el paso de los años.

El DL es un tipo de ML que emplea una estructura artificial neuronal jerarquizada que sigue un diseño similar al del cerebro humano con nodos neuronales que se conectan entre sí como si fuera una tela de araña o una red de neuronas del tejido cerebral humano. Estos nodos de la *red neuronal artificial* se separan en distintas capas de espesor que se conectan entre sí. Encontramos una primera capa de entrada, una serie de capas intermedias conocidas como capas ocultas (de procesamiento) y una última capa de salida (de clasificación) (**Figura 1.35**). En el entrenamiento del algoritmo se emplean datos de los que se conoce su resultado final; por ejemplo, una preparación digitalizada que se conoce que es un cáncer agresivo. Las distintas capas de la red realizan las siguientes funciones:

- En la *capa de entrada* cada nodo recibe un único dato específico de la muestra (por ejemplo el tamaño del núcleo), que se procesa, se extraen sus características principales y se le asigna un peso numérico según su importancia antes de ser transferido a la siguiente capa.

- En la *capa oculta*, cada nodo usa métodos estadísticos y matemáticos que tienen en cuenta el peso o valor asignado previamente para clasificar el dato. Si este dato transformado alcanza un valor concreto, se activa la neurona y es enviada a los siguientes nodos; y así sucesivamente según el número de capas que presente la red.

- En la *capa de salida* se realiza la predicción y se clasifica finalmente el dato inicial (maligno o benigno) (Marcial García Rojo, 2021). El resultado que se

> predice se compara con el resultado verdadero previamente conocido, de manera que el sistema realiza un análisis inverso y aprende qué características o factores ayudan a obtener un resultado correcto; y puede eliminar las conexiones y neuronas que lo lleven a una clasificación incorrecta. Este procedimiento en el que solo se mantienen los resultados aceptables se conoce como "poda de neuronas", según Le CUN (L. Wang, 2017).

Este es el método de funcionamiento del DL, que engloba numerosos algoritmos, entre los que destacan las redes CNN para la clasificación y reconocimiento de imágenes. Las redes CNN están ganando popularidad por ser fundamentales para tratar de resolver problemas sobre imágenes, con innovadoras aplicaciones (Bernal et al., 2019) ofreciendo muy buenos resultados en la clasificación de imágenes (Ahmad et al., 2017) .

Los sistemas de ML presentan algunas desventajas cuando las comparamos con los sistemas modernos de DL. Los algoritmos de ML tienen la principal desventaja de que pueden sufrir el fenómeno de *sobreentrenamiento*, que consiste en exceder el punto óptimo de entrenamiento, a partir del cual comienzan a empeorar los resultados. Esto no ocurre en los modelos de DL, que continúan mejorando los resultados conforme se van entrenando, sin llegar a saturarse. Entre las ventajas del uso de la técnica del DL encontramos que no es necesario el procesamiento previo de las imágenes y se puede trabajar sobre las imágenes en bruto directamente. Como inconveniente, se encuentra el requerimiento de una gran base de datos para manejar un volumen de datos suficiente para realizar un buen entrenamiento del sistema, y que el DL requiere un tiempo de entrenamiento mayor que los ML por manejar muchos más parámetros, aunque el tiempo empleado en la realización de las pruebas es menor.

1.6. LA APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL EN ANATOMÍA PATOLÓGICA

1.6.1. DIGITALIZACIÓN E INICIOS DE LA PATOLOGÍA DIGITAL

La digitalización en Anatomía Patológica ha permitido el desarrollo e integración de la IA en la patología digital (PD). La digitalización de muestras mediante escáneres especializados, permite obtener imágenes digitales de las muestras de una calidad idéntica o superior a la obtenida con su visionado al microscopio. A través de visores digitales, se puede "navegar" por la muestra digital, emitir un diagnóstico, realizar fotografías y compartir los casos con otros profesionales a distancia, mediante la *telepatología*. En la actualidad la PD es una de las apuestas más fuertes para la modernización y mejora de la calidad diagnóstica en Anatomía Patológica (Parra et al., 2020), (Bueno et al., 2016).

La digitalización tiene numerosas ventajas, como la disminución de los tiempos de respuesta y la optimización de los flujos de trabajo. La posibilidad de digitalizar las muestras de los pacientes permite disponer de grandes fuentes de datos y su intercambio con otros profesionales, mejorando la eficiencia y potenciando la colaboración intra e interdepartamental, reduciendo así los errores médicos (Niazi et al., 2019). Además, constituye una herramienta fundamental en la docencia, ya que permite realizar sesiones clínicas, realizar cursos o seminarios "online", disponiendo de un gran banco de muestras digitales (Parra et al., 2020), (Rojo, 2019).

El término *patología digital,* se utilizó por primera vez a principios de los 2000, refiriéndose al uso de imágenes completamente digitalizadas obtenidas mediante el escaneo de portaobjetos, generando imágenes digitales de toda la muestra de pacientes, conocido en inglés como *whole slide digital imaging* (Saltz, 2000), (Rojo, 2019). Este término pasó a sustituir a otros sinónimos anteriores como patología virtual o microscopía virtual (Weinstein et al., 2009). En la actualidad el término PD es más amplio, ya que engloba a todas aquellas herramientas de comunicación e informáticas que ayudan al proceso diagnóstico; tanto en la gestión digital de datos

como en la gestión de imágenes, con el uso de escáneres y de sistemas de visionado y almacenamiento (Rojo, 2019), (Acs et al., 2020).

En los inicios de la PD empezaron a elaborarse herramientas de análisis automatizado de imágenes para cuantificar biomarcadores o para la segmentación de áreas de interés, que no llegaron a ser muy populares entre los patólogos ya que obligaban a dar el paso a la digitalización, en un periodo en el que el uso de la PD era muy escaso en los laboratorios de Anatomía Patológica. Además, los escáneres del momento no eran de gran calidad y los resultados de estos análisis automatizados dependían de la manera directa de la calidad y resolución de la imagen obtenida (Rojo, 2019).

La digitalización presenta una gran complejidad en el manejo de los datos, requiriendo hasta 5 gigabytes de almacenamiento por cada imagen digitalizada (Acs et al., 2020). Con el avance de la medicina, a estos datos se les suma el almacenamiento de cada vez más información clínica del paciente (como datos moleculares o genómicos) (Rojo, 2019). Según García Rojo, ante esta necesidad de almacenamiento de un gran volumen de datos; tanto de gestión de datos (clínicos, radiológicos, farmacológicos, etc.) como de imágenes digitalizadas, surgió en el año 2010 el concepto *patología computacional*, que hace referencia al empleo de la IA en el tratamiento complejo con métodos probabilísticos de flujos de trabajo en Anatomía Patológica (Marcial García Rojo, 2021), (Fuchs & Buhmann, 2011), (Rojo, 2019). La patología computacional busca elaborar herramientas de apoyo diagnóstico y de toma de decisiones que ayuden al patólogo; trabajando tanto con la imagen microscópica como con datos moleculares o genómicos, mejorando la precisión y eficiencia diagnóstica (Althammer et al., 2019).

1.6.2. LOS ESCÁNERES EN PATOLOGÍA DIGITAL

Los escáneres de portaobjetos son una herramienta fundamental en PD, que han ido evolucionando rápidamente con los años. Según García Rojo (Marcial García Rojo, 2021), un escáner puede escanear a alta resolución (40x) unas 500 preparaciones un día de trabajo (a 0,25 micras/pixel).

Los escáneres de carga continua admiten la recarga con bandejas nuevas para el escaneo sin la necesidad de parar la maquinaria del escaneado. Se pueden escanear sobre la marcha nuevos casos, incluyendo casos urgentes y posponiendo los menos relevantes.

Los escáneres que no son de carga continua tienen un número máximo de cristales que se pueden escanear a la vez, generalmente 6, lo que limita mucho el flujo de trabajo y requiere una buena planificación del trabajo. Una vez iniciado el proceso, no se puede añadir ningún caso más hasta que finalice (Parra et al., 2020).

En los escáneres se pueden personalizar distintos parámetros como el aumento (10x, 20x o 40x), la elección del número de niveles o la selección de un área de escaneado. A mayor área de escaneado, número de capas o aumento; mayor tiempo de escaneado y peso del archivo resultante (Parra et al., 2020). Para un resultado óptimo, en biopsias se recomienda un escaneado a 40x y en citología un escaneado de al menos 3 niveles (Thorstenson et al., 2014), preferentemente de citologías líquidas ya que son más limpias y de mejor calidad.

Por ejemplo, una biopsia pequeña cutánea escaneada a 40x pesará entre 400-700 megabytes, frente a una biopsia grande cutánea que pesará entre 3-4 gigabytes (GB). Existen muestras digitales grandes de hasta 40 GB que pueden comprimirse en ficheros de 1-3 GB (Marcial García Rojo, 2021). Un laboratorio de Anatomía Patológica completamente digitalizado, puede generar 500 portaobjetos por jornada laboral, con una media de 2 GB cada uno, lo que implicaría 1 terabyte de datos comprimidos al día (Marcial García Rojo, 2021).

1.6.3. HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL DE AYUDA AL DIAGNÓSTICO EN ANATOMÍA PATOLÓGICA

El *gold estándar* en el diagnóstico de enfermedades es la morfología microscópica de los tejidos. Sin embargo, presenta una importante limitación, la variabilidad diagnóstica entre patólogos en patologías poco frecuentes o de gran dificultad (Chang et al., 2019). La necesidad de dar diagnósticos de precisión en la patología tumoral está aumentando a medida que van desarrollándose distintos biomarcadores que determinan el tratamiento del cáncer. Sin embargo, numerosos países del mundo se enfrentan a un problema importante: la escasez de patólogos (Acs et al., 2020). Es por ello que la IA se está convirtiendo en una herramienta de apoyo al diagnóstico muy prometedora en los servicios de Anatomía Patológica, que se está convirtiendo en un campo de acción ideal para DL (Rojo, 2019).

La aplicación del DL sobre una imagen digital trata de buscar patrones que permitan a la red diferenciar qué es positivo (tejido tumoral) o negativo (tejido sano). El DL aplicado a Anatomía Patológica trabaja con preparaciones digitales que el algoritmo descompone en distintos fragmentos o teselas. El sistema DL recibe datos con los que trabajar (un portaobjetos digitalizado) y una tarea a resolver (discernir si es positivo o negativo) de manera autónoma (Chang et al., 2019). Según Miller y cols., existen otras muchas herramientas de ayuda diagnóstica, como el procesamiento de datos para que se generen mapas de calor que marquen las zonas de interés, como las áreas positivas para el tumor (Miller & Brown, 2018).

La digitalización y el análisis de imágenes promete mejorar la precisión en la evaluación histológica de tejidos y marcadores moleculares. El desarrollo del DL ha permitido grandes avances en la patología computacional y marcará el centro su integración con la atención sanitaria en la próxima década, donde la histopatología estará en el centro (Acs et al., 2020). Se ha mostrado que estas aplicaciones de ML ofrecen mejores resultados que el examen humano en la detección de metástasis de células malignas en el ganglio linfático; en la cuantificación del índice proliferativo con ki67 en el cáncer de mama; en la

clasificación de Gleason en el cáncer de próstata; y la evaluación de la presencia de linfocitos intraepiteliales en el melanoma. Algunos modelos de DL son capaces de predecir el estado de los biomarcadores antes de su determinación en cánceres como el de pulmón, colon, próstata o gástrico, basándose únicamente en la morfología de las células con la tinción estándar con hematoxilina-eosina. También se han desarrollado modelos de redes CNN que; a partir de la morfología de las células, son capaces de ofrecer un pronóstico y cifras de supervivencia de algunas neoplasias como el cáncer de pulmón, melanoma o glioma (Acs et al., 2020).

Las ventajas del uso de la IA en los servicios de Anatomía Patológica son superponibles a los encontrados en el uso de la PD, como la mejora y optimización de los flujos de trabajo, la automatización de algunas tareas de diagnóstico (como la cuantificación de los biomarcadores de algunos tumores), la mejora de situaciones de carencia de patólogos o citotécnicos a la hora del diagnóstico y la mejora en la eficiencia de los profesionales («Libro Blanco de la Anatomía Patológica en España 2021», 2021). Uno de los resultados más llamativos que se están obteniendo con el uso de la IA es la mejora de la reproductibilidad en diagnósticos que clásicamente presentan gran variabilidad interobservador debido a su dificultad, además de la disminución de los errores y la cuantificación precisa parámetros pronósticos (B et al., 2020). Según Moxley-Wyles y cols., se reduciría un 40% el tiempo de respuesta en los servicios de Anatomía Patológica si se incorporaran a los laboratorios la IA junto a un flujo de trabajo digital (B et al., 2020).

Por lo tanto, ante las aplicaciones de la IA en Anatomía Patológica encontramos:

- Herramienta como cribado rápido de muestras. Dentro de la carga de trabajo del patólogo, el algoritmo es capaz de localizar qué preparaciones digitales son relevantes (como las tumorales), para priorizarlas sobre el resto de muestra. Este tipo de herramientas ahorrarían la revisión de hasta el 75% de preparaciones que no contienen alteraciones relevantes, sin disminuir la sensibilidad de la técnica (Campanella et al., 2019), (B et al., 2020). Por ejemplo, en la búsqueda de metástasis en los ganglios en el cáncer

de mama, se estima una disminución del 58,2% de la carga del trabajo para el patólogo (Acs et al., 2020).

- Mejora en la concordancia del diagnóstico entre patólogos (Rakha et al., 2021), ayuda en la determinación del grado de agresividad o clasificación tumoral (Acs et al., 2020) y como herramienta de apoyo para no pasar por alto ningún área importante en la preparación que afecte al diagnóstico (B et al., 2020).

- Herramienta de segunda opinión diagnóstica y de telepatología. Gracias al empleo de algoritmos de IA en Radiología, según Moxley-Wyles y cols. se ha disminuido la necesidad de una segunda opinión en el 88% de los casos (B et al., 2020).

- Posibilidad de relacionar las características de la muestra con el riesgo de recurrencias, de mutaciones génicas o supervivencia del paciente, estableciendo la IA una predicción del pronóstico (Acs et al., 2020).

- Apoyo en docencia, investigación y control de calidad (Rakha et al., 2021).

Del mismo modo, los programas de Ayuda al Diagnóstico Médico asistidos por Computador (*Computer Assited Diagnóstic systems o CAD*), están ganando interés debido a que ofrecen ventajas sobre el análisis humano experto. Superan el cansancio físico y psíquico humano que desencadena una tarea prolongada en el tiempo, que puede afectar a los resultados (García-Laencina et al., 2014) (García-Laencina et al., 2015). Otra ventaja de los CAD es que pueden aplicarse como herramienta de ayuda instantánea, realizando una valoración inicial in situ del paciente, como si se tratara de la valoración del caso por un Comité de expertos. Esta ayuda diagnóstica como segunda opinión tiene un gran valor en varios escenarios; como en pequeños centros clínicos donde acceder a una segunda opinión es compleja o en hospitales de alta carga asistencial en los que duplicar los esfuerzos en llegar al diagnóstico preciso puede suponer un problema organizativo.

1.6.4. TÉCNICAS DE MACHINE LEARNING EN PATOLOGÍA

Existen tres formas de aprendizaje en el campo del ML (Chang et al., 2019). En el ML con *aprendizaje supervisado*, el algoritmo entrena sobre imágenes digitales con varias anotaciones de las zonas de interés marcadas (por ejemplo, las áreas con tumor), que el ser humano ha interpretado previamente manualmente o de forma semiautomática. En el caso del *aprendizaje no supervisado*, el algoritmo emplea un sistema de datos de entrenamiento sobre los que aprende a distinguir patrones sin la acción del ser humano. El *aprendizaje por refuerzo* la red toma decisiones interactuando con el entorno a través de recompensas o penalizaciones, formando un sistema de ensayo y error.

Se ha demostrado que las técnicas "supervisadas" son superiores a las "no supervisadas" y cuanto mayor es el número de imágenes con las que se alimenta el algoritmo, se obtienen mejores resultados. Sin embargo, el rendimiento del proceso de anotación de imágenes por el patólogo es bajo. Se estima que en una jornada laboral completa, un patólogo no anotará manualmente más de 10 imágenes digitalizadas, según Rojo y cols. (Rojo, 2019).

Es por ello, que en la práctica se trata de emplear una acción intermedia (*ML débilmente supervisados* o anotados) donde se clasifican grandes sistemas de datos que han sido previamente diagnosticados por un ser humano, pero que no tienen anotaciones detalladas sobre las zonas de más interés (Robertson et al., 2018). Tiene la ventaja de evitar anotaciones costosas que consumen mucho tiempo en píxeles. Campanella y cols., realizaron pruebas con este sistema sobre el cáncer de próstata, carcinoma de células basales y metástasis ganglionar del cáncer de mama; dando unos resultados de área bajo la curva mayores a 0,98 y establecieron que su aplicación permitía liberar a los patólogos de entre el 65-75% de los portaobjetos, manteniendo una sensibilidad del 100% (Campanella et al., 2019).

Una de las soluciones que se aplica para no tener que anotar de manera manual miles de anotaciones, es el uso de la *segmentación automática o smart segmentation*, que son redes CNN en las que el algoritmo anota las zonas de interés

de la imagen y el patólogo se limita a mejorar los resultados y corregir los posibles errores. Otra opción es el empleo del *data augmentation* que consiste en emplear imágenes originales y modificarlas (mediante rotaciones, traslaciones, cambios en el contraste, en la escala de la imagen, etc.) para crear imágenes nuevas. Otros métodos son el *aprendizaje discriminativo* que requiere menor número de imágenes (Vu et al., 2016) o el *transfer learning* que aprovecha redes entrenadas con imágenes de otras categorías (incluso ajenas al mundo médico), sobre las que solo hay que entrenar las últimas capas de la red profunda, disminuyendo el número de imágenes necesarias para entrenar al algoritmo (Ker et al., 2019).

1.6.5. REDES NEURONALES CONVOLUCIONALES APLICADAS AL DIAGNÓSTICO DE PREPARACIONES DIGITALES

Las CNN han experimentado un gran éxito en la clasificación de imágenes médicas, de las que se están obteniendo resultados muy prometedores con el paso de los años (Goodfellow I et al., 2016). En el diagnóstico de imágenes celulares, Godinez y cols., aplicaron una CNN multiescala para estadificar distintos fenotipos celulares a través de imágenes, siendo los resultados obtenidos mediante CNN más precisos que los obtenidos con otros sistemas (Godinez et al., 2017). En 2016 Moeskops y cols., presentaron un método de segmentación automática de resonancias magnéticas nucleares de tejido cerebral, utilizando una resonancia como entrada y obteniendo segmentaciones muy acertadas (Moeskops et al., 2016). Sun y cols., publicaron un trabajo de CNN multicanal para segmentar tumores de hígado a través de imágenes de tomografía computarizada con contraste en multifase mejorada, obteniendo datos mejorados de las características patológicas en cada fase (Sun et al., 2017). En 2017 se presentó una CNN con la capacidad de detectar melanomas a través de imágenes dermatoscópicas de lunares con una exactitud superponible a la de dermatólogos expertos, según Esteva y cols. (Esteva et al., 2017). Otra aplicación de CNN según Motlagh y cols., es la clasificación de tipos y subtipos de cáncer de mama empleando laminillas histopatológicas teñidas con hematoxilina-eosina y preparaciones inmunohistoquímicas (Motlagh et al., 2018).

Hay buenos resultados en el empleo de CNN en la clasificación de imágenes en Anatomía Patológica; tanto de citologías como de biopsias (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013). Las CNN son las redes más utilizadas en el análisis de imágenes de biopsias y citologías en Anatomía Patológica (LeCun et al., 2010), (Chang et al., 2019). Cada neurona de CNN calcula un valor de salida aplicando un vector de pesos y sesgos a los valores de entrada procedentes de la capa anterior (Janowczyk & Madabhushi, 2016). El entrenamiento u optimización del modelo CNN proviene de ajustar estos sesgos para minimizar la pérdida, considerando que cuantos más datos se introduzcan, mayor rendimiento se obtendrá. Para evaluar la eficacia de los sistemas DL supervisados, se pueden comparar los resultados obtenidos por el algoritmo con la clasificación de los patólogos expertos y aplicar parámetros estadísticos como: sensibilidad, especificidad, precisión, valor predictivo positivo o negativo.

Las publicaciones sobre el empleo de las CNN en el diagnóstico de biopsias en Anatomía Patológica se centran en los cánceres más comunes (mama, próstata o pulmón) ya que el modelo requiere un gran entrenamiento. Sin embargo, se están obteniendo buenos resultados en otros cánceres menos frecuentes como el melanoma (Hekler et al., 2019), tumores neuroendocrinos pancreáticos (Niazi et al., 2019), cáncer de ovario (Wu et al., 2018), cáncer de cuello de útero (L. Zhang et al., 2017) y el glioma (Ertosun & Rubin, 2015).

El diagnóstico del cáncer de mama mediante CNN se inició con el análisis de imágenes citológicas de punciones e imágenes de tejido teñidas con hematoxilina-eosina, según Ehteshami y cols., (Ehteshami Bejnordi et al., 2018). Se trabajó con 175 imágenes capturadas por una cámara microscópica a 40x de aumento y se dividieron en 918 regiones de interés (ROI) de 256 x 256 píxeles de tamaño, donde cada ROI incluía múltiples células (Garud et al., 2017). Se entrenó al algoritmo para evaluar la malignidad de una región de interés y se determinó que la muestra era maligna cuando >30% de las regiones de interés eran clasificadas como malignas, obteniendo grandes resultados (Chang et al., 2019).

El uso de la CNN en la clasificación de células en el cérvix uterino mediante la citología cervicovaginal, fue explorado por Zhang y cols., que realizaron un procesamiento previo de las imágenes en el que eliminaron cualquier artefacto que dificultara el visionado de la célula, con las que entrenaron a la red. El inconveniente de este tipo de estudios es que emplean; tanto en el entrenamiento como en la evaluación, células aisladas sin solapamientos ni artefactos, las cuales no son la situación real habitual de diagnóstico. En una citología normal, las células escamosas generalmente se presentan en contacto con otra diversidad de células epiteliales e inflamatorias, junto a flora bacteriana saprofita, solapamientos y agrupaciones celulares (L. Zhang et al., 2017).

De esta forma la CNN se presenta como una herramienta muy prometedora en la clasificación automática de citologías cervicovaginales aplicada a programas de cribado poblacional. Sin embargo, según las revisiones de Alves y cols., del *Scottish Cervical Cytology Review Group* (2009), del *Medical Services Advisory C.* (2009) y de Queiro Verdes y cols., no hay ningún sistema en la actualidad suficientemente fiable para implantarlo en el cribado (Alves et al., 2004), (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013),.

A pesar de presentarse como una herramienta de precisión en la clasificación celular con una tasa de precisión en ocasiones mayor a la del patólogo, presentan una gran desventaja: su escasa explicabilidad (Martínez-Más et al., 2020), (Taha et al., 2017). Son técnicas que se consideran cajas negras, donde no se tiene la información de por qué la red toma su decisión. Este rasgo genera mucha desconfianza entre los profesionales sanitarios, debido a que; el no entender los motivos de la decisión del algoritmo, dificulta seriamente la implantación de estos sistemas DL en la asistencia sanitaria.

1.6.6. RETOS Y BARRERAS EN LA APLICACIÓN DE LA IA EN ANATOMIA PATOLÓGICA

El uso de la IA está implantado en algunas especialidades médicas en las que se ha demostrado una gran utilidad en el desempeño del trabajo diario. Los avances más prometedores en patología computacional se basan en la IA y el DL, que ofrecen modelos informáticos de reconocimiento de imágenes que igualan o superan a los humanos expertos (Acs et al., 2020).

A pesar de que están ampliamente reconocidas las ventajas del uso de la IA en Anatomía Patológica y de que existen numerosas pruebas que respaldan que la colaboración hombre-máquina mejora el rendimiento de ambas partes (Miller & Brown, 2018), (Niazi et al., 2019), existen algunas limitaciones en la implantación de estos sistemas. En Anatomía Patológica el análisis computacional de imágenes existe desde hace muchos años (Bengtsson, 1987), (Ong et al., 1996) pero su aplicación en la patología rutinaria está fuertemente limitado por retos y barreras, según Niazi (Niazi et al., 2019):

- El uso de imágenes de gran tamaño, que pueden alcanzar hasta los 100.000 x 100.000 píxeles. Esto marca una gran limitación en la digitalización de portaobjetos que requieren programas informáticos especializados, largos tiempos de procesamiento y la necesidad de un espacio de almacenamiento muy amplio para almacenar gran cantidad de datos (Acs et al., 2020).

- El uso de imágenes complejas con gran variedad de colores dependiendo de las técnicas tintoriales que se empleen (hematoxilina-eosina, histoquímica o inmunohistoquímica), con ausencia de orientación anatómica y que requieren combinar la información obtenida de varios aumentos (ya que cada aumento aporta una información diferente y complementaria).

- La imposibilidad de disponer de un sistema de patología digital completo (B et al., 2020) y la necesidad de un gran nivel de computación para entrenar a los algoritmos de Deep learning (Robertson et al., 2018).

- La necesidad de trabajar con un volumen de imágenes anotadas alto para que el algoritmo ofrezca buenos resultados, cuando en la realidad en Anatomía Patológica el volumen de imágenes con anotaciones es muy bajo (Campanella et al., 2019), (Niazi et al., 2019).

En la actualidad se han desarrollado herramientas que tratan de resolver algunas de las barreas que se van encontrando:

1. Imágenes artificiales. Algunos algoritmos de redes neuronales son capaces de generar imágenes nuevas a partir de un conjunto de imágenes del mismo tipo. El algoritmo aprende la tinción, características de la imagen y su forma; y las reproduce en imágenes artificiales que se confunden con imágenes auténticas, incluso por patólogos expertos (Acs et al., 2020). Este sistema permite aumentar el número de imágenes disponibles para entrenar al algoritmo.

2. Transferencia del aprendizaje (*transfer learning*). Consiste en reutilizar algoritmos que fueron entrenados para otras tareas; por ejemplo para localizar coches en fotografías, y se adaptan para el desarrollo de otras tareas, como identificar áreas tumorales. Es muy útil el empleo de ImageNet, una base de datos con millones de imágenes de distintos objetos (Rojo, 2019), (Robertson et al., 2018).

3. Aumento de imágenes mediante la transformación de imágenes disponibles. Utilizando imágenes etiquetadas, se pueden realizar cambios pequeños que no modifican su naturaleza pero que sirven para que se considere una imagen distinta de la original, consiguiendo así multiplicar el número de imágenes disponibles (Rojo, 2019), (Robertson et al., 2018). Se pueden realizar cambios como cambio de color, desenfoque, escalado, rotaciones, volteos, deformaciones y recortes.

4. Métodos de anotación de imágenes de fácil aplicación. Estos métodos se basan en un *aprendizaje débilmente supervisado*, en el que cada preparación presenta una etiqueta asociada de diagnóstico que permite una clasificación automática de las imágenes (Robertson et al., 2018). También hay diseñados algoritmos que tratan de facilitar una anotación automática de imágenes digitales (Niazi et al., 2019). Respecto a la anotación de imágenes hay dos tendencias: (1) entrenar al algoritmo con imágenes perfectamente depuradas con anotaciones perfectas y (2) entrenar al algoritmo con un gran número de imágenes que incluyan todos los artefactos posibles que se dan en condiciones reales, para que el algoritmo aprenda a distinguirlos (Acs et al., 2020).

1.7. LA IA EN CITOPATOLOGÍA

1.7.1. FUNDAMENTO Y BASES DEL DESARROLLO DE LA IA EN CITODIAGNÓSTICO

Existen sistemas de análisis automatizado aplicado al diagnóstico de citologías cervicovaginales que; según Alves y cols., no han demostrado mejores resultados (de sensibilidad ni especificidad) que el análisis humano para el diagnóstico de lesiones CIN 2 o superiores, independientemente del tipo de citología utilizada (convencional o líquida) o el sistema de análisis automático empleado (Alves et al., 2004). Sin embargo, en temas de rendimiento o productividad, evaluados mediante la medición de horas de lectura y número de muestras diagnosticadas por día, se ha visto que la lectura automatizada es superior al diagnóstico manual (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013).

Según Jiang y cols., la aplicación de la lectura automatizada en el cribado citológico del cáncer de cérvix mejora la eficacia general del cribado y reduce los costos de atención médica derivados la detección y tratamiento del cáncer, sobre todo en estadios avanzados (Jiang et al., 2023). Sin embargo; al comparar los

análisis coste-efectividad de los sistemas automatizados aplicados al cribado poblacional con el examen manual, vemos que los sistemas automatizados son más caros, según las revisiones publicadas en el *Medical Services Advisory C*. (2009) (Kitchener et al., 2011) y el *Scottish Cytology Review Group (2009).* Según Arbyn y cols., ninguno de estos sistemas conocidos de DL aplicados al cribado poblacional mediante citología, está recomendado por el Programa Europeo contra el Cáncer, debido a la falta de evidencia necesaria para la elaboración de guías clínicas (Arbyn et al., 2010). Se debe a que la identificación de la displasia puede ser laboriosa y puede tener baja consistencia diagnóstica por la alta similitud entre algunas de las imágenes patológicas de SIL (An et al., 2023).

El empleo del análisis automatizado presenta una mayor eficiencia que el análisis manual, ya que se analizan más muestras por unidad de tiempo (Kitchener et al., 2011), (Boost, 2009). Se estima que se necesitan entre 5 y 10 minutos para examinar una citología, dependiendo de la dificultad de la muestra y la experiencia del citotecnólogo. Debido a la creciente demanda de citotecnólogos y patólogos por el gran volumen de citologías que se realiza como método de cribado, el examen manual puede ser tedioso y está sujeto a errores. El desarrollo de sistemas de detección automatizados del cáncer de cérvix está en apogeo y está destacando como herramienta que permite optimizar los flujos de trabajo (Jiang et al., 2023).

En los estudios de Kitchener y cols., y Boost y cols., (Kitchener et al., 2011), (Boost, 2009) emplean la red sobre la citología líquida, que selecciona los 20 campos con mayor atypia, para que sean revisados de manera manual. Al reducir el tiempo de visionado mediante la selección de las áreas sospechosas, se ahorra tiempo y se optimiza el trabajo de visionado, obviando las citologías que son normales, según Queiro Verdes y cols. (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013). Además, según Jiang y cols., el estudio de toda la muestra hace necesario ir desarrollando sistemas de selección de áreas de interés para segmentar células y clasificarlas como precancerosas o no (Jiang et al., 2023).

1.7.2. ANTECEDENTES DEL EMPLEO DE CNN COMO MODELO DE CLASIFICACIÓN CELULAR EN CITOLOGÍAS

Los métodos tradicionales ML tratan de segmentar el contorno de la célula para así poder extraer las características del núcleo y citoplasma, cuyas alteraciones definen la displasia. Por lo tanto, la *segmentación celular* es clave para la extracción de características, que puede verse dificultada el la práctica clínica real por artefactos en la imagen, como superposición celular o el fondo sucio que borran los límites celulares. Aplicado a los modelos de CNN, se introduce como entrada la imagen segmentada directamente al modelo para la extracción de sus características citológicas y la obtención de una clasificación en una categoría diagnóstica en su capa de salida.

En general, el éxito del DL se relaciona directamente con la cantidad de datos introducidos para realizar el entrenamiento. Una cantidad de datos insuficiente empeora el rendimiento del modelo, como ocurre en ocasiones en el análisis de imágenes médicas con falta de anotaciones o etiquetas (Jiang et al., 2023). Según la revisión sistemática de Jiang y cols., (Jiang et al., 2023) de todo lo publicado desde el año 2016 hasta la actualidad, los modelos de CNN emplean tres enfoques de entrenamiento para evitar esta problemática: el *aprendizaje por transferencia*, el *conjunto multimodelo* y la *fusión de características híbridas* (**Anexo 1**).

A. APRENDIZAJE POR TRANSFERENCIA

El aprendizaje por transferencia reutiliza un modelo entrenado previamente en una tarea específica para emplearla en una nueva tarea distinta. Se toma un modelo entrenado con un conjunto de datos (por ejemplo, de ImageNet) y se adapta a su nueva tarea mediante ajustes: (1) se congelan las primeras capas de la CNN (no se modifican porque contienen características básicas comunes a cualquier imagen como bordes, texturas o patrones de colores) y (2) se ajustan las últimas capas de la red (especializadas en la nueva tarea del modelo). Por último,

(3) se realiza un nuevo entrenamiento del modelo con nuevos datos específicos para desempeñar su nueva tarea.

Un ejemplo del aprendizaje por transferencia, es el empleo de una red preentrenada en reconocer animales del zoológico para una nueva tarea de distinguir razas de perro. Al reutilizar la red, se pueden aprovechar las capas que reconocen las características generales de los animales (tamaño, forma, pelaje, color, etc.) y solo sería necesario ajustar las últimas capas específicas en el reconocimiento de la raza de perro.

Estos modelos tienen ventajas y inconvenientes, sobre todo si se emplean en tareas especializadas como las imágenes médicas. Las ventajas principales del empleo de aprendizaje por transferencia son:

(1) La disminución del tiempo de entrenamiento ahorrando tiempo de etiquetado y entrenamiento de la red desde cero.

(2) La necesidad de menor cantidad de datos para el entrenamiento de la red en la nueva tarea (ya que por su antiguo entrenamiento sabe reconoce características básicas útiles).

(3) Un mejor rendimiento inicial al comienzo del nuevo entrenamiento debido a que ya tiene la capacidad previa de reconocer imágenes.

Respecto a los inconvenientes encontramos:

(1) El riesgo de sobreajuste (*overfitting*, en inglés) ya que si el nuevo conjunto de datos es pequeño y no se ajustan bien las últimas capas el modelo memoriza estrictamente los datos nuevos sin ser capaz de generalizar al enfrentarse a imágenes nuevas.

(2) La necesidad de ajuste fino (en inglés, *fine-tuning*), que consiste en la necesidad de hacer ajustes adicionales de todas las capas, para lo que puede ser necesario una gran cantidad de datos y de tiempo, sobre todo si el nuevo dominio es muy distinto del modelo original.

(3) La pérdida de la interpretabilidad de los patrones aprendidos en el modelo preentrenado, al aplicarle la nueva tarea. Esto dificulta la tarea de comprender cómo el modelo toma sus decisiones, especialmente relevante en campos como la medicina.

(4) Desajuste entre la tarea original y la nueva tarea, debido a que el modelo preentrenado puede haber sido entrenado para la detección de imágenes muy distintas a las de su nuevo entrenamiento. Aplicado al campo de las imágenes médicas, puede observarse este desajuste ya que tras el entrenamiento previo del modelo, podría aprender a extraer patrones que son irrelevantes para esta nueva tarea.

(5) Ante una tarea muy especializada; como en el diagnóstico de imágenes médicas, el modelo preentrenado podría no aportar características útiles sin un entrenamiento exhaustivo, pudiendo ser más efectivo y eficiente entrenar la red desde cero aunque requiera de más tiempo y datos.

De este modo, este tipo de aprendizaje se emplea con frecuencia cuando se trabaja con imágenes médicas debido a que; la obtención de grandes cantidades de datos etiquetados y el tiempo necesario para entrenar la red desde cero, no siempre está disponible en medicina, ya que estas imágenes suelen ser difíciles de obtener y etiquetar por expertos. También se ha demostrado una mayor eficiencia en el empleo de tiempo y recursos al reutilizar el modelo preentrenado.

Esta técnica ya ha mostrado éxito en la clasificación de imágenes médicas en áreas como la detección de cáncer, la segmentación de órganos en imágenes de resonancia magnética, la detección de retinopatía diabética y la clasificación de tumores. Sin embargo, puede tener algunas limitaciones ya que las imágenes

médicas difieren mucho de las imágenes generales con las que se suele entrenar la red, siendo más abstractas y sutiles con texturas, sombras y detalles que pueden ser claves para el diagnóstico. Es por ello que es fundamental un ajuste minucioso y fino de las capas o incluso el reemplazo completo de capas.

Aplicado a la citología cervicovaginal, DeepPap es un ejemplo de transferencia de imágenes con grandes resultados (L. Zhang et al., 2017), que emplea el modelo ConvNet simple preentrenado con imágenes naturales de ImageNet, que luego se ajustó con datos citológicos cervicales. Se trata de una red neuronal que permite clasificar células de citologías cervicovaginales en 7 clases distintas, que abarcan 6 tipos de células escamosas (células normales superficiales, intermedias, displasia leve, moderada, grave y carcinoma epidermoide in situ) y 1 tipo de células columnares. Se observó con este sistema una sensibilidad de 98,2% y especificidad del 98,3%.

Posteriormente se han publicado un conjunto de imágenes médicas de células de citologías cervicovaginales, como SIPaKMeD o Herlev, mejorando la precisión de los modelos que se han ido desarrollando posteriormente.

B. CONJUNTO MULTIMODELO

El aprendizaje conjunto multimodelo aplicado a los modelos CNN, consiste en el empleo de varios modelos que trabajan juntos para resolver una tarea. En lugar de depender de una sola CNN, se combinan o fusionan las predicciones de múltiples modelos para mejorar la precisión y la robustez de los resultados. Este enfoque puede ser especialmente útil en tareas donde un solo modelo podría no ser lo suficientemente confiable o donde la variabilidad de los datos es alta.

En este tipo de modelos; en primer lugar se debe realizar un entrenamiento independiente de los distintos modelos CNN, para después combinar sus salidas de datos para producir una única predicción final: ponderando las predicciones, o

promediando las predicciones mediante el uso de la media o la selección del voto mayoritario de todas las salidas.

En los últimos años, el DL conjunto multimodelo se ha aplicado ampliamente en los campos biomédicos y bioinformáticos (Ganaie et al., 2022). Al combinar varios modelos se reduce el riesgo de que uno cometa errores críticos mejorando la precisión general multimodelo, siendo más robusto, ya que es capaz de captar aspectos de los datos que un único modelo podría pasar por alto.

Por ejemplo, si empleamos este sistema en el diagnóstico médico de imágenes obtenidas por resonancia magnética para detectar cáncer, se emplearían 3 modelos distintos que detectan distintas características. El modelo 1 podría detectar elementos básicos como la forma general del tumor, el modelo 2 podría centrarse en detalles complejos como la textura del tejido en la imagen y el modelo 3 se trataría de un modelo entrenado previamente con imágenes médicas aportando datos generales. Una vez diseñadas se llevarían a cabo las predicciones por separado de cada modelo, en el que cada uno detallaría el porcentaje de probabilidad de que la imagen contenga un cáncer. Por último, se elaboraría una predicción final global, generalmente empleando la media aritmética del resultado de los tres modelos. Este sistema de combinación de modelos hace que el diagnóstico sea más preciso y fiable.

Este sistema tiene ventajas e inconvenientes. Entre las ventajas encontramos:

(1) Mayor precisión y rendimiento en la predicción final. Cada modelo empleado aporta una perspectiva diferente, como detalles finos, detalles generales o aportando experiencia previa con imágenes.

(2) Reducción de los errores ya que si un modelo comete un error, los otros compensarán, ya que no todos los modelos fallarán en la misma imagen.

(3) Mayor robustez del sistema frente a variaciones en los datos. Diferentes modelos pueden aprender diferentes características de los datos y, combinados, pueden proporcionar una visión más completa y precisa.

(4) Disminución del sobreajuste, debido a que los modelos individuales pueden llegar a sobreajustarse a los datos del entrenamiento capturando características muy específicas de la imagen y presentando importantes dificultades para generalizar ante nuevos datos. Al combinar las predicciones de varios modelos, el conjunto es capaz de generalizar mejor.

Las desventajas del sistema multimodelo son:

(1) Mayor gasto computacional ya que al entrenar a varias CNN requiere más recursos en tiempo y en hardware.

(2) Mayor complejidad en la implementación debido a que se deben combinar las salidas de los distintos modelos y gestionar la mayor carga computacional.

(3) Dificultad en la interpretación ya que; ante múltiples modelos trabajando juntos, puede ser más difícil entender qué características están siendo captadas por cada modelo y cómo contribuyen a la predicción final. Esto puede ser un problema en áreas donde la interpretabilidad es crucial, como la medicina.

En el campo de las imágenes médicas, los conjuntos multimodelo han sido utilizados para tareas como la clasificación de enfermedades en imágenes de rayos X, resonancias magnéticas o tomografías computarizadas. Diferentes CNNs pueden aprender a detectar diferentes tipos de anomalías o características, y al combinar sus salidas, se puede mejorar la precisión del diagnóstico.

El empleo de este sistema en la identificación de células cervicales ha mostrado grandes resultados. En 2021 Rahaman y cols. propusieron un modelo que empleaba 4 redes DL (VGG16, VGG19, XceptionNet y ResNet50) para la

clasificación multiclase de células cervicales, elaborando una predicción final con una precisión de hasta 99,85% para dos clases (Rahaman et al., 2021). En el mismo año, Diniz y cols. propusieron un multimodelo empleando los datos públicos CRIC para el entrenamiento (CRIC cervix collection) que combinaba 3 modelos obteniendo muy buenos resultados (Diniz et al., 2021).

C. FUSIÓN DE CARACTERÍSTICAS HÍBRIDAS

Aunque los modelos de DL han mostrado grandes resultados en la clasificación de células cervicales, hay rango de mejora. Existen características celulares muy específicas, sutiles, que presentan gran relevancia en el diagnóstico del cáncer. La incorporación de conocimientos del ámbito médico a la red de DL puede potenciar la eficacia de la red mejorando su rendimiento.

La fusión de características híbridas, de forma global, consiste en la combinación de distintas características o tipos de información para obtener mejores resultados. Trata de juntar las características de varias fuentes de información para mejorar la precisión del modelo, combinando características heterogéneas (características de diferentes tipos o niveles) en lugar de depender de un único conjunto de características extraídas por la red. Aplicado a los modelos CNN se refiere a la combinación de diferentes tipos de características extraídas de distintas fuentes de datos para mejorar el rendimiento de la red en una tarea específica, como la clasificación o la segmentación de imágenes. Es decir, en las CNN tradicionales las características extraídas suelen ser únicamente visuales y se obtienen a través de las capas convolucionales, que identifican patrones como bordes, texturas, formas y otros detalles visuales. Sin embargo, en muchos problemas del mundo real, combinar estas características visuales con otro tipo de información (como características a diferentes niveles de abstracción, textos, tablas o incluso de diferentes tipos de datos) puede mejorar el rendimiento de la red.

El empleo de estos modelos híbridos se usa mucho en aplicaciones complejas como el diagnóstico médico o el reconocimiento de objetos, ya que permite al modelo tener una mejor idea de lo que está analizando. Esta fusión de características híbridas puede llevarse a cabo de distintas formas: (1) mediante la combinación de características a distintas profundidades de la red para obtener una representación más completa de la imagen; (2) con la fusión de distintas fuentes de datos que aportan información adicional a las imágenes (como datos, textos o tablas); (3) la combinación de características de distintos modelos CNN (como modelos que capturan características temporales y otros espaciales) y (4) la fusión de características manuales extraídas por expertos y automáticas detectadas por la red.

Según el tipo de fusión que se emplea en el modelo distinguimos: los *métodos de fusión temprana* cuando se fusionan las características antes del inicio del modelo CNN; *métodos de fusión intermedia* cuando las características se combinan a medida que se van generando; *métodos de fusión tardía* cuando se combinan justo antes de realizar la predicción final y la *fusión jerárquica* cuando combina las características de múltiples niveles de la red de manera jerárquica. La fusión jerárquica tiene la ventaja del aprovechamiento tanto de características locales y finas (bajas) como de las globales y abstractas (altas), de manera que se logra representar de forma completa los datos.

Por ejemplo, al aplicarlo al campo de diagnóstico de imágenes médicas, podríamos emplear una red CNN para diagnosticar enfermedades a través de radiografías de tórax. Esta fusión de características híbridas podríamos realizarla combinando características de distintos niveles, como las características simples de las primeras capas (bordes o texturas de la imagen) con las características más complejas de las últimas (detección de manchas patológicas o anomalías), esperando que con la combinación de detalles finos con la visión general, mejore el diagnóstico. Otro modo sería usando distintas fuentes de información, combinado la radiografía con otros datos relevantes como la historia clínica del paciente. Por último, podríamos combinar distintos tipos de modelos; por ejemplo combinando una CNN que detecte detalles a pequeña escala (como pequeñas manchas en los

pulmones), y otra que se enfoque en estructuras más grandes (como la forma completa de los pulmones).

Este sistema presenta las siguientes ventajas:

(1) Mejoría del rendimiento y precisión del modelo al tener una visión más completa y al lograr una mejor representación de los datos por combinar características de distintas fuentes o niveles. Estas características permiten mejorar el rendimiento de la red en tareas de clasificación, segmentación o detección.

(2) Permite combinar información diversa, no limitando la tarea al análisis de imágenes, debido a que su combinación con otros datos como texto o tablas de datos, puede ayudar a hacer predicciones más precisas.

(3) Mejor generalización. Al utilizar múltiples fuentes o tipos de características, el modelo tiene más capacidad para generalizar a nuevos datos, ya que puede basar su predicción en una mayor variedad de información.

(4) Mayor robustez del modelo. Al fusionar características de diferentes tipos, el modelo puede ser más robusto frente a variaciones en los datos de entrada, ya que se basa en información complementaria de varias fuentes o niveles.

Los inconvenientes que podemos encontrar son los siguientes:

(1) Complejidad computacional al aumentar la cantidad de datos que deben ser procesados, aumentando la necesidad de memoria y los tiempos de procesamiento.

(2) Dificultad en la integración de las distintas características provenientes de niveles.

(3) Riesgo de sobreajuste ante los datos de entrenamiento, sobre todo si el conjunto de datos es pequeño o las características no son realmente útiles para la tarea.

(4) Dificultad en la interpretabilidad del modelo al combinar distintos tipos de características. Esto conlleva un riesgo de no entender las bases de la predicción para confiar en los resultados.

En la detección de enfermedades con imágenes médicas, la fusión de características híbridas es especialmente útil para combinar las características visuales de la imagen con otra información adicional, como los antecedentes clínicos del paciente. A su vez, para mejorar el rendimiento de la red, la introducción de más información nueva en modelos DL más allá de los datos médicos existentes da un enfoque prometedor (X. Xie et al., 2021).

Los citopatólogos expertos pueden aprovechar su experiencia para mejorar los modelos de DL en la clasificación de las células cervicales. Una manera es mediante la integración de anotaciones manuales con las características extraídas por el modelo, ya que las características hechas a mano reflejan un amplio conocimiento biomédico en la morfología celular. Gu y cols., emplearon CE-Net para segmentar las células cervicales y se añadieron a mano características morfológicas, densidades ópticas y textura para una clasificación más precisa (Gu et al., 2019). También realizaron una regularización suavizada de etiquetas ruidosas y una estrategia de aprendizaje contrastivo supervisado para entrenar al modelo. Se logró una precisión del 98,96% y del 98,67%, respectivamente.

En el año 2020, Dong y cols., propusieron un algoritmo de DL que combina características hechas a mano con las funciones extraídas automáticamente de la red Inception v3 (Dong et al., 2020). Emplearon una capa que fusionaba ambas características y emplearon un algoritmo de mejora de imágenes para reducir el ruido generado durante la adquisición y conversión de imágenes para mejorar el rendimiento general. Este método propuesto consiguió una precisión del 98,34% en la clasificación en dos clases y del 94,68% en la clasificación en siete clases.

En 2024 Sha y cols. diseñaron un modelo híbrido llamado CerviFusionNet que combina redes CNN con transformadores de visión y módulos de memoria recurrente, trabajando con imágenes básicas, informes clínicos, diagnósticos patológicos e imágenes clínicas de pruebas con ácido acético, obteniendo resultados superiores a los métodos convencionales (Sha et al., 2024).

1.7.3. SISTEMAS DE DETECCIÓN DISPONIBLES EN LA ACTUALIDAD

De manera simultánea al desarrollo de los sistemas de cribado automático de primera generación, las técnicas de elaboración y procesamiento de las muestras han evolucionado constantemente. El desarrollo de la citología líquida ha permitido a los ordenadores localizar y visualizar células individuales de manera más fácil debido a la distribución uniforme de células, facilitando enormemente el desarrollo de herramientas de diagnóstico automatizado de citologías cervicovaginales.

Con el desarrollo de la IA se han ido elaborando sistemas de diagnóstico automatizado de citologías cervicovaginales, algunos de ellos aprobados para su comercialización en España. La aprobación de la FDA certifica una evaluación dela seguidad y eficacia por los Estados Unidos. Por el contrario, el marcado CE certifica que organismos acreditados de la Unión Europea corroboran el cumplimiento de las normas esenciales de rendimiento y seguridad.

La FDA aprobó en 2004 y 2008 los productos comerciales *ThinPrep Imaging System®* (Hologic, Bedford, MA, USA) y *FocalPoint GS Imaging System* ® (BD Diagnostics-TriPath, Burlington, NC. USA), que se aplicaban sobre sus propios sistemas de citología líquida comercializados *ThinPrep®* y *AutoCyte®* (Surepath), respectivamente (Biscotti et al. 2005), (Kardos 2004). En 2024, la FDA aprobó el sistema de Hologic, *Genius™ Digital Diagnostics System.* Además, estos tres sistemas también cuentan con el marcado CE.

Por otro lado, otros sistemas como *SureView™ Cervical Cytology System* (Techcyte) y *BestCyte® Cell Sorter Imaging System* (CellSolutions) presentan únicamente el marcado CE.

ThinPrep Imaging System®

ThinPrep Imaging System® se aplica sobre citologías líquidas teñidas con el método de tinción comercializado *ThinPrep®* de *Hologic Imager®* (Koss et al., 1994). Este sistema consta de dos partes, el *Image Processor* que prepara la muestra eliminando todos los artefactos o material no diagnóstico que pueda alterar el visionado (homogeneizando las células y separándolas de la sangre, moco o detritus) y el *Review Scope* que se encarga del visionado del portaobjetos (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013). Para ello se escanea todo el portaobjetos, el sistema lo revisa y selecciona los 22 campos visuales con mayor probabilidad de presentar anomalías celulares para analizarlos con mayor detenimiento. Estos campos se presentan al *Review Scope*, que se trata de un sistema de microscopio con platina motorizada permite revisar los 22 campos a estudio por un citotécnico o un patólogo (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013). Si el sistema encuentra alguna alteración celular en alguno de los campos seleccionados, toda la muestra debe ser revisada. En ningún caso se asigna a la muestra una categoría diagnóstica, pero tiene la ventaja de que se le evita al patólogo revisar muestra catalogadas como normales, sin ningún campo alterado.

En algunos hospitales tras la implantación de este sistema, se ha evidenciado una mejora de la precisión y eficiencia en el diagnóstico citológico, mostrando un aumento estadísticamente significativo de los diagnósticos de LSIL y HSIL al compararlos con el diagnóstico manual (Chivukula et al., 2007). Se ha obtenido una sensibilidad en la detección de ASC-US del 66%, LSIL del 88% y HSIL del 88% (Crowell et al., 2019). A pesar de que presenta cifras de sensibilidad altas (94,3%), no emplea ningún sistema de diagnóstico celular, sino que actúa como un método de asistencia por imagen que selecciona las áreas de mayor riesgo (Akca & Erkilic, 2023).

BD FocalPoint GS Image System

BD FocalPoint GS Image System es el nombre que recibe el antiguo sistema de primera generación *AutoPap Primary Screening System* desde el año 2001 (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013), que puede emplearse tanto en citología convencional como en citología líquida tras emplear el sistema de tinción comercializado por ellos, el *SurePath Pap Test*. Tras la toma de la muestra, se eliminan los elementos celulares no diagnósticos como moco, sangre, detritus y células inflamatorias para optimizar el material, se elabora la citología y se digitaliza. Este sistema consta de dos componentes, el *Slide Profiler* y el *Review Station.* El *Slide Prolifer* clasifica las citologías en "revisión" y "no revisión adicional" siempre limitando esta segunda clasificación a un 25% del volumen de citologías a estudio. Las catalogadas como "revisión" requieren una revisión manual y las "no revisión adicional" se catalogan como citologías normales. A estas citologías seleccionadas como "revisión" se le aporta un valor numérico en quintiles dependiendo de la probabilidad de contener células patológicas (quintil 1 mayor riesgo y quintil 5 menor riesgo). Cuando se observa un quintil de 1-15 hay que revisar la citología por lo que estas muestras se remiten al *Review Station,* que selecciona los 10 campos con mayor probabilidad de presentar anomalías mediante un microscopio de platina motorizada, para ser revisados manualmente (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013).

Algunos estudios demuestran una sensibilidad de 81,1-86,1% y especificidad de 84,5-85,1% (Gupta et al., 2023) para el diagnóstico de citologías cervicovaginales con este sistema, a pesar de tratarse de un algoritmo clásico no CNN que se centra en priorizar y clasificar pacientes y señalar las áreas de interés en la citología.

GeniusTM Digital Diagnostics System

Este sistema de diagnóstico digital cuenta tanto con la aprobación de la FDA como con el CE. Se digitaliza todo el portaobjetos para obtener una WSI para que posteriormente el algoritmo seleccione las áreas interés para exportarlas a una

estación de revisión con un monitor. De este modo ofrece una galería de 30 regiones o células sospechosas que el citotecnólogo o patólogo debe revisar para emitir un diagnóstico final. Al igual que los sistemas anteriores, no emite un diagnóstico final automatizado, centrándose en priorizar las zonas sospechosas mediante IA para la revisión por un experto.

Este tipo de sistemas destaca por agilizar los tiempos de trabajo, ya que el experto en lugar de revisar toda la citología, únicamente examina las 30 áreas seleccionadas por la IA. Sin embargo, al contrario que otros algoritmos, no sirve para cribar y descartar a las pacientes sanas ya que no emite diagnósticos, por lo que el experto debe revisar las imágenes seleccionadas de todas las citologías. Se han publicado diferencias de tiempo de diagnóstico muy significativas, con 3,2 minutos en el diagnóstico digital asistido por la red y 5,9 min con el microscopio (Cantley et al., 2024). Este sistema presenta una gran precisión y detecta un 7,5% más de casos HSIL que el examen convencional de la citología al microscopio y presenta unas cifras de sensibilidad del 98,2-100% (Elishaev et al., 2025).

SureViewTM Cervical Cytology System

Este sistema emplea una red CNN que analiza y ordena las células según sus características para su revisión posterior. Para ello se digitaliza la citología (elaborada con ThinPrep o SurePath) para que el algoritmo CNN busque y clasifique distintas regiones celulares de interés, para pasarlas a un visor para su revisión posterior por un patólogo o citotecnólogo experto que confirme o descarte la presencia de patología en esas células. Este sistema está validado para optimizar los flujos de trabajo y para actuar como un sistema de control de calidad.

Ha demostrado una sensibilidad del 82%, especificidad del 99%, precisión del 97%, valor predictivo negativo del 98% y valor predictivo positivo del 86% (Rivera Rolon et al., 2025). Sin embargo, presenta el fuerte inconveniente de que solo es compatible con 5 escáneres disponibles en el mercado y requiere de una gran infraestructura y sistemas informáticos para su implantación.

BestCyte

Es un sistema que emplea algoritmos de clasificación tras la digitalización de todo el portaobjetos, que ordena en imágenes las células para una revisión. Gracias a su red CNN es capaz de identificar distintas características celulares y repartir las imágenes en distintas galerías según los hallazgos observados y ordenarlas entre sí según importancia, como: *atypical, halos, elongated, clusters, high N/C,* entre otros. Posteriormente el patólogo experto emitirá un diagnóstico tras la revisión de las imágenes.

Este sistema ha demostrado una sensibilidad del 100% y una especificidad del 93,3% en el diagnóstico de células atípicas (Chantziantoniou, 2023). Sin embargo la especificidad; o capacidad de detectar células benignas, es menor situándose en un 67,08%.

Existen varios estudios que tratan de evaluar la seguridad, efectividad relativa y coste efectividad al comparar la lectura automatizada con estos sistemas aprobados por la FDA al compararlos con la lectura manual. Se ha visto un mayor rendimiento y mayor productividad en la lectura automatizada frente a la manual al comparar factores como duración de la lectura y muestras analizadas por día (Teresa Queiro Verdes & Jeannette Puñal Riobóo, 2013). Por otro lado, respecto al análisis coste-efectividad observado en los estudios Australianos MSAC y de Reino Unido MAVARIC, la lectura automatizada de citología líquida presenta menor coste-efectividad al compararla con la lectura manual líquida o convencional, dentro del contexto de programas de cribado poblacionales del cáncer de cérvix, por lo que estos autores no recomiendan la implantación del sistema automatizado en el cribado poblacional.

1.7.4. FASES EN LA IMPLANTACIÓN DE UN DIAGNÓSTICO AUTOMATIZADO

Para implantar un sistema de diagnóstico automatizado hay que coordinar una serie pasos para garantizar unos buenos resultados. Jiang y cols. plantean en su revisión sistemática los pasos necesarios para la implantación de un diagnóstico automatizado, que se desarrollan en el **Anexo 2**. (Jiang et al., 2023). Desarrollan las estrategias que distintos grupos de trabajo han empleado en la literatura para mejorar los resultados de la red, utilizando distintos enfoques de trabajo.

Según los trabajos publicados, la fase de *adquisición y preprocesamiento de imágenes* es fundamental para obtener imágenes de alta calidad. Bai y cols., emplearon modelos de tinción virtuales basados en DL para obtener imágenes de gran calidad, logrando así omitir los pasos de tinción convencionales y liberando al personal de laboratorio y la infraestructura requerida para emplearla en otras tareas (B. Bai et al., 2023). También se han empleado redes GAN para unificar el estilo de las imágenes escaneadas, ya que con frecuencia se observa cierta variabilidad en la tinción de las citologías una vez escaneadas (Creswell et al., 2018). Croitoru y cols., emplearon modelos generativos profundos basados en GAN y modelos de difusión para mejorar las imágenes con baja resolución o desenfocadas (Croitoru et al., 2023).

En la fase de *análisis de la imagen* encontramos distintos métodos según la bibliografía consultada. En general, se han establecido las redes DL como el mejor método para obtener mejores resultados (Simonyan & Zisserman, 2015), (Szegedy et al., 2015), (S. Xie et al., 2017), (K. He et al., 2016). Se ha visto que si se emplea una red demasiado profunda, se obtienen peores resultados ya que se requiere un mayor volumen de datos para su entrenamiento. En los últimos años Niu y cols., y Guo y cols., (Niu et al., 2021), (Guo et al., 2022) han propuesto un sistema de diagnóstico a través de un *mecanismo de atención* que trata de imitar el funcionamiento del cerebro humano, que escanea la información global y centra la mirada (en el centro de la retina) en un campo de visión concreto para obtener

información local más cuidadosamente. De esta forma se centra en la información importante y elimina la información ruidosa no relevante.

1.7.5. RETOS DEL CITODIAGNÓSTICO AUTOMATIZADO

A pesar de que se han logrado grandes avances en la clasificación automatizada de las citologías cervicovaginales en los últimos años, todavía se presentan desafíos considerables que se intentan superar.

Para obtener un buen rendimiento de los sistemas DL y CNN es importante introducir una gran cantidad de información. Existen plataformas de entrenamiento previo de modelos de DL mediante imágenes naturales como ImageNet; sin embargo, el conjunto de datos de imágenes de alta calidad para citología cervicovaginal es limitado. Las razones recaen en el alto costo en la recopilación de datos, ya que se requiere pasar por un proceso de preparación del portaobjetos, tinción y escaneado que resultan costosos en términos de equipamiento y mano de obra. Además, debido a que el etiquetado es un proceso tedioso que requiere experiencia por parte del patólogo, solo una parte de estas muestras tiene anotaciones con la etiqueta de diagnóstico que incluya la ubicación de la célula problema. Es compleja la obtención de suficientes casos patológicos con un volumen equilibrado de células de todas las categorías del Sistema Bethesda para lograr un conjunto de datos equilibrado de entrenamiento.

El empleo de distintos estilos de imagen supone también un gran desafío. La imágenes digitales obtenidas de las citologías para el entrenamiento de los algoritmos, generalmente presentan distintos estilos, ya que se pueden ver condicionados por el tipo de tinción citológica empleada, la calidad de procesamiento y los distintos escáneres de portaobjetos empleados. Esta heterogeneidad en los estilos de imagen entre los datos de entrenamiento y los de prueba, da lugar a un bajo rendimiento de los modelos entrenados, dificultando la construcción de modelos DL sólidos y generalizados. Por ello, han cobrado fuerza distintos métodos de normalización de la tinción de las imágenes; tanto

tradicionales (con transferencia del color, coincidencia espectral de manchas, etc), como mediante métodos basados en DL que utilizan redes generativas adversarias (Chen et al., 2021).

Otra barrera radica en que no siempre se tiene acceso a un escaneo de calidad de las citologías cervicovaginales. La técnica de superresolución de imágenes, trata de convertir diapositivas de baja resolución en alta resolución, ya que las imágenes de desenfocadas pueden interferir en el diagnóstico preciso de citologías cervicovaginales (Ma et al., 2022). Esto se presenta como una gran oportunidad para aquellos territorios subdesarrollados que no pueden adquirir escáneres avanzados.

Hay otras vías de investigación que tratan de corregir la superposición celular. En la práctica es muy frecuente la presencia de grupos de células superpuestas que dificultan el diagnóstico y hay varios autores que tratan de solventar esta limitación, incluso con técnicas de DL (Tareef, Song, Huang, et al., 2017), (Tareef, Song, Cai, et al., 2017).

Sin embargo, en base a nuestra experiencia, este tipo de imágenes preprocesadas son poco realistas al compararlas con las analizadas por un patólogo experto con un microscopio óptico. Tras el procesamiento y tinción de la citología es muy frecuente que las células presenten artefactos, solapamiento, plegamiento o incluso superposición con otros tipos celulares como células inflamatorias o microorganismos. Por lo tanto, si mejoramos en exceso la calidad de la imagen en la fase de preprocesamiento, estamos falseando la calidad de las características morfológicas celulares que serían visibles en condiciones normales.

II – JUSTIFICACIÓN

II - JUSTIFICACIÓN

Para los servicios de Anatomía Patológica, la citología cervicovaginal acarrea un gran peso del total de citologías que se elaboran y se examinan día a día. Estas tomas citológicas no solo se realizan en las consultas de ginecología; sino que también se obtienen en los Centros de Atención Primaria como método de cribado del cáncer de cérvix, lo que aumenta en nivel de carga sociosanitaria que estas representan. En los hospitales del área 2 de Salud del Servicio Murciano de Salud (Hospital Santa María del Rosell y el Hospital General Universitario Santa Lucía de Cartagena) ha aumentado el registro de citologías con el paso de los años (**Figura 2.1**):

- Entre los años 2015 a 2020, se realizaron un total de 38.055 citologías (que incluyen citologías de orina, cervicovaginales, lavados bronquiales, aspirados bronquiales y muestras de líquido cefalorraquídeo), de las cuales el 53% fueron exclusivamente citologías cervicovaginales (un total de 20.021). Además, según los datos oficiales, en 2019 se realizaron 12.600 citologías en los centros de atención primaria, que por convenio se analizaron en centros externalizados (con una media anual de unas 13.000 citologías). Por lo tanto, se estima que en los últimos 5 años se han realizado aproximadamente en el Área 2 de Salud; juntando las muestras de atención hospitalaria y de atención primaria, unas 103.055 citologías.
- Con la implantación de la citología cervicovaginal como método de cribado en los Centros de Salud del Área 2 en 2021, el volumen de diagnóstico de citologías cervicovaginales (hospitalarias y de cribado en atención primaria) se ha incrementado un 175,2% en los últimos 5 años.
- Entre el año 2021 y el 2025 el volumen de las citologías cervicovaginales (incrementado por el sobreañadido de muestras correspondientes al cribado de los centros de salud) representa el 73% de todas las citologías estudiadas en nuestro Área de Salud (**Figura 2.2**).

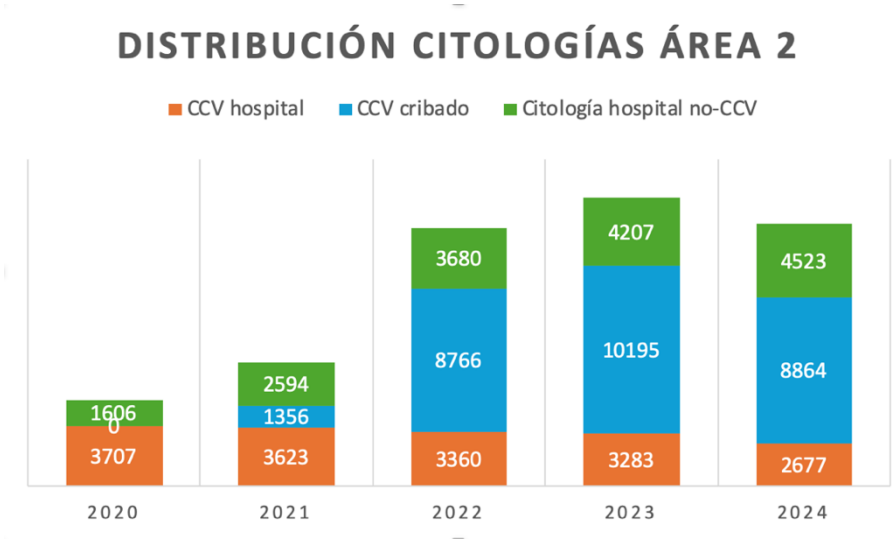


Figura 2.1. Distribución de las citologías en nuestro complejo hospitalario entre los años 2020-2024, repartido entre citologías cervicovaginales hospitalarias ("CCV hospital"), citologías cervicovaginales del cribado de atención primaria ("CCV cribado") y el resto de citologías de origen no ginecológico obtenidas en el centro hospitalario ("citología hospital no-CCV"). Podemos ver cómo todos los años las citologías cervicovaginales representan más de la mitad de la carga de trabajo del Área 2. Elaboración propia.

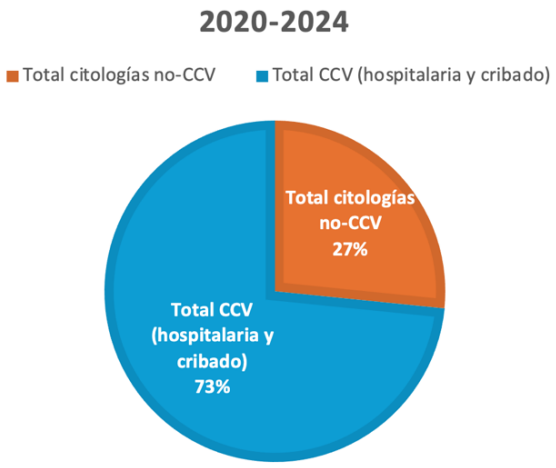


Figura 2.2. Flujo de trabajo del diagnóstico citológico en el Área 2 de Salud del 2020-2024. En conjunto durante 5 años las citologías cervicovaginales (hospitalarias y del cribado en atención primaria) suman un 73% del volumen total de muestras citológicas. Elaboración propia.

Estos datos ponen de manifiesto la gran importancia que tienen este tipo de muestras; no solo por su relevancia en el diagnóstico de la patología cervical, sino también por el gran volumen de trabajo que representa para los servicios de Anatomía Patológica. Por esta razón con el paso de los años se intenta simplificar el proceso de la toma de muestras (con el paso de la triple toma a la citología líquida); agilizar la elaboración de las citologías en el laboratorio (con el empleo de maquinaria de procesamiento y tinción automatizados) y protocolizar el diagnóstico con criterios estandarizados (con la aplicación de guías diagnósticas internacionales como el Sistema Bethesda de citología cervicovaginal (Ronco et al., 2007) (Arbyn et al., 2008). Sin embargo; aunque se intente agilizar la cadena diagnóstica, esta siempre finaliza con la visión manual y subjetiva del patólogo, cuya precisión dependerá de la fatiga del observador y de posibles errores del laboratorio, entre otros. Otra nueva herramienta diagnóstica que se está implementando con los años y con la que se está obteniendo muy buenos resultados en otros campos de la medicina, es el empleo de la IA como herramienta de apoyo al diagnóstico para intentar mejorar la sensibilidad de este tipo de pruebas, intentando superar a los resultados obtenidos por expertos.

Con el desarrollo de la patología digital, se introduce el concepto de digitalización de las muestras en Anatomía Patológica («Libro Blanco de la Anatomía Patológica en España 2021», 2021). A través de escáneres especializados, se escanean las muestras a alta definición para que a través de monitores de alta resolución se pueda trabajar con "muestras digitales". La digitalización tiene innumerables ventajas entre las que destacan:

- La posibilidad de realizar una teleconsulta que brinda la posibilidad de poder compartir en tiempo real imágenes de casos clínicos reales para; en consenso con otros profesionales sanitarios que están a distancia, poder llegar al diagnóstico preciso de la enfermedad del paciente.
- La facilidad de almacenamiento de bancos de imágenes digitales en la red sin tener que disponer de almacenes especializados en los centros hospitalarios.

- La disponibilidad de acceder a estas imágenes desde cualquier ordenador, que representa una herramienta imprescindible para la docencia hospitalaria mediante la elaboración de bibliotecas digitales o la elaboración de talleres de diagnóstico online.
- La posibilidad de trabajo remoto o teletrabajo para el patólogo, que ya no depende del uso del microscopio para trabajar, disminuyendo la posibilidad de extravío de muestras o portaobjetos cuando se trasladan casos clínicos entre centros hospitalarios cuando es necesario.

La implantación de la patología digital en el trabajo asistencial se está extendiendo poco a poco y tiene cada vez mayor aceptación entre los servicios de Anatomía Patológica. Sin embargo, la automatización completa en la lectura de muestras no está aceptada y hoy día el examen de las muestras lo sigue realizando un patólogo.

Este proyecto plantea desarrollar un algoritmo de DL (en inglés, *Deep Learning*) que permita etiquetar las muestras de citologías cervicovaginales en cada una de las categorías diagnósticas del Sistema Bethesda, participando en el diagnóstico como haría un patólogo experto. Para ello se trata de segmentar las muestras previamente digitalizadas en miles fotografías de células independientes para que el algoritmo trate de clasificar cada una de ellas en una de las distintas categorías diagnósticas para así emitir un diagnóstico final, analizando variables de la imagen como el color, forma o textura de la célula. La base teórica que sustenta esta propuesta es la aplicación de los criterios diagnósticos del Sistema Bethesda de manera similar a como lo haría el patólogo en el diagnóstico de la patología cervicovaginal. Por ejemplo y de forma general, criterios como núcleos de pequeño tamaño, de perfil liso y redondeado suelen pertenecer a células benignas; mientras que los núcleos agrandados, muy teñidos de un color oscuro intenso y de perfil irregular suelen pertenecer a lesiones premalignas, en ocasiones de alto. De la misma manera, tras un entrenamiento adecuado, el algoritmo debería poder ser capaz de analizar dichas variables para emitir un diagnóstico final.

En estudios preliminares (Martínez-Más et al., 2020) se ha planteado como problemática principal la escasez de células de algunas de las categorías diagnósticas, ya que la distribución del número de células por categoría no es homogénea en las citologías en condiciones normales y esto dificulta el entrenamiento del algoritmo. Esta problemática se tratará de resolver en este estudio y buscaremos elaborar un modelo con una precisión diagnóstica de calidad al compararla con la obtenida por patólogos experimentado.

La idea de este proyecto no solo va enfocada como un apoyo para mejorar la carga asistencial del patólogo y así reducir fatigabilidad, costes y tiempos de diagnóstico; sino también para su uso como herramienta de ayuda diagnóstica en casos ambiguos que pueden presentar una gran variabilidad interobservador. Según un estudio comparativo publicado en Noruega sobre las diferencias en los diagnósticos de HSIL (CIN 2) obtenidos por cuatro patólogos; se vio que la sensibilidad variaba entre el 68.8% y 93.8% y la especificidad entre 70.6% y 95.6% (Sørbye et al., 2017) entre los distintos patólogos del estudio.

Por lo tanto, este algoritmo podrá testarse como ayuda para el patólogo en el diagnóstico de las citologías; no solo como apoyo en la tarea mejorando la sensibilidad, especificidad y disminuyendo errores de la técnica, sino también agilizando todo el proceso diagnóstico y reduciendo así el impacto de la carga asistencial sociosanitaria.

III – OBJETIVOS

III - OBJETIVOS

3.1. OBJETIVO PRINCIPAL

- OP.- Desarrollar un algoritmo Deep Learning para etiquetar imágenes de células en cada una de las categorías diagnósticas de la citología cervicovaginal como una herramienta de ayuda al diagnóstico del cáncer de cérvix.

3.2. OBJETIVOS ESPECÍFICOS

- OE1.- Evaluar la precisión diagnóstica del algoritmo comparándolo con los diagnósticos emitidos por patólogos expertos.

- OE2.- Desarrollar un modelo hibrido que combine una red generativa adversaria y una red CNN para la creación de células cervicales artificiales, evaluando su utilidad y fiabilidad como herramienta para ampliar nuestra base de datos.

- OE3.- Explorar métodos de interpretabilidad con la segmentación celular como estrategia de diagnóstico por imagen frente a modelos de clasificación convencionales.

- OE4.- Implementar el mapeo por activación de clases ponderado por gradiente (Grad-CAM), como método de explicabilidad para mejorar la confianza de los profesionales sanitarios al modelo.

- OE5.- Adaptar el algoritmo de diagnóstico celular para entornos de recursos limitados con baja disponibilidad a equipos de digitalización adecuados.

IV - MATERIAL Y MÉTODO

IV -MATERIAL Y MÉTODO

El proyecto ha seguido varias líneas investigación que se detallan en distintas secciones. En este capítulo se desarrolla en primer lugar la base de datos empleada para el desarrollo del proyecto, que es común para todas las líneas de investigación. Más adelante veremos como se ha trabajado la base de datos desde distintos aspectos o líneas de investigación para la mejora de los resultados del algoritmo, que se desarrollarán en el apartado de resultados.

Este proyecto cuenta con la aprobación del Comité de Ética de la Universidad Católica de Murcia, con el código CE012005, con fecha 27 de julio 2021.

4.1. BASE DE DATOS

4.1.1. BASE DE DATOS AL INICIO DEL PROYECTO

Se ha realizado una búsqueda del archivo de citologías cervicovaginales de pacientes del Área 2 de Salud obtenidas mediante el sistema de citología líquida *BD SurePath® liquid Pap test®* (**Figura 4.1-A**). Estas citologías fueron realizadas con anterioridad al inicio del estudio, y en su día fueron procesadas de forma habitual en el servicio de Anatomía Patológica del Complejo Hospitalario Universitario de Cartagena mediante el sistema automatizado *PrepStain®* y *SuperPap®* (**Figura 4.1-B**). Cada citología va ligada a un diagnóstico realizado por un patólogo de entre las siguientes categorías del Sistema Bethesda: (1) citología normal (células benignas), (2) lesión escamosa intraepitelial de bajo grado (LSIL), (3) lesión intraepitelial

escamosa de alto grado (HSIL) y (4) atipia de células escamosas de significado incierto (ASC-US).

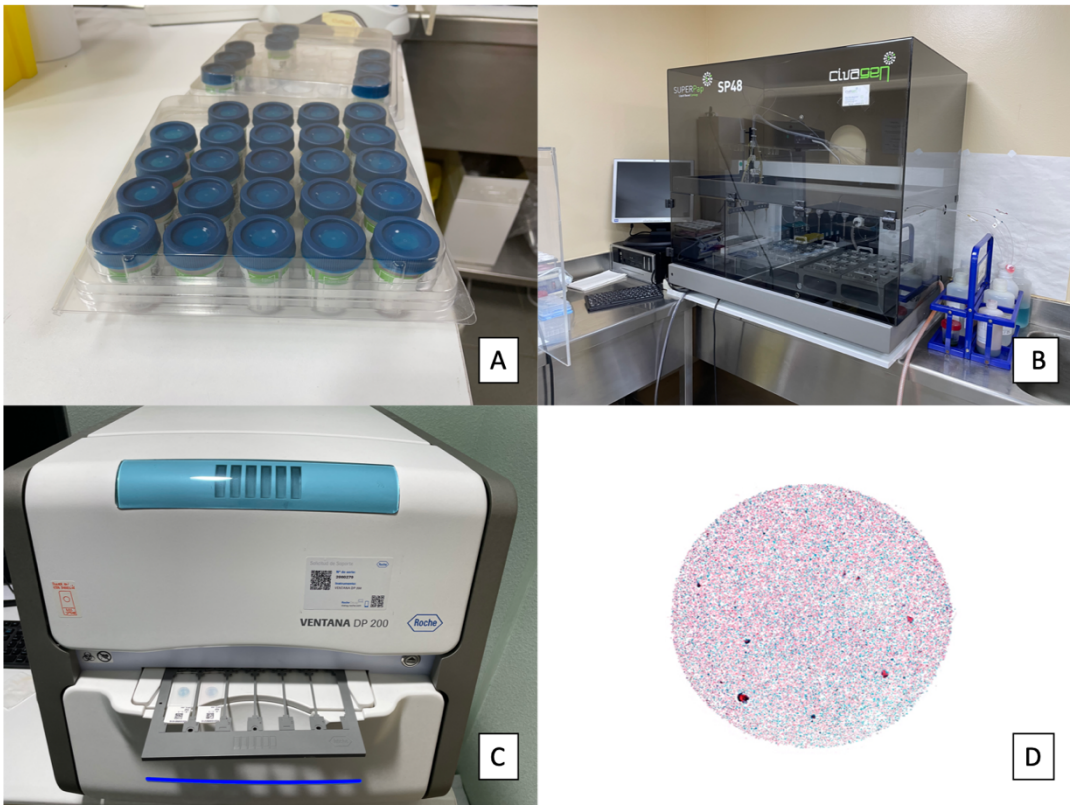


Figura 4.1. Flujo de trabajo para la elaboración y escaneado de una citología cervicovaginal. A.- Sistema de recogida de la muestra del paciente mediante un cepillado cervical que se remite a un fijador celular BD SurePath® liquid Pap test®. B.- Procesado automático citológico mediante el sistema SuperPap®. C.- Escáner Ventana DP200 de escaneado de laminillas. D.- Imagen digital obtenida de todo el portaobjetos (whole slide image, WSI) sobre la que podemos navegar y modificar su aumento para el estudio celular a alta definición. Elaboración propia.

Una vez seleccionado este volumen de citologías representativas de cada una de las categorías diagnósticas del Sistema Bethesda, fueron digitalizados los portaobjetos mediante el escáner Ventana DP200 (Roche®) (**Figura 4.1-C**), un escáner de laminillas que permite el escaneo de un máximo de 6 cristales a la vez con unos aumentos 20x y 40x en un formato .TIFF o .jpeg200. También se ha

empleado el escáner de carga continua EasyScan (Motic®) que es más rápido y escanea a alta definición en formatos .jpg, .svs y .mds.

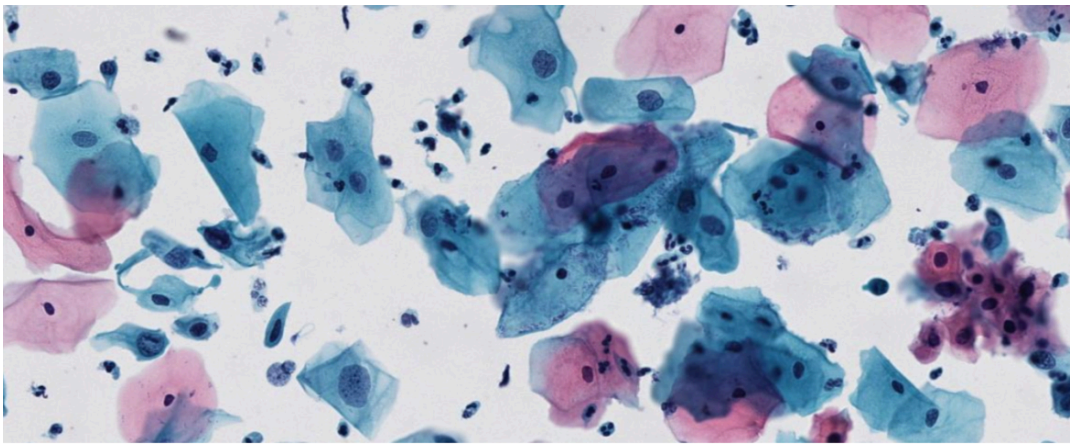


Figura 4.2. Imagen digital de una citología cervicovaginal escaneada con el escáner Ventana DP200 a un aumento digital 20x. Elaboración propia.

Tras el escaneado, se ha obtenido una la imagen digital de todo el portaobjetos (en inglés *whole slide image*, WSI) (**Figura 4.1-D**), que mediante distintos softwares especializados se puede visionar en un ordenador, navegando sobre la imagen, tomando fotografías y aumentando o disminuyendo el aumento de la célula observada al igual que se hace con un microscopio (**Figura 4.2**).

Posteriormente, se dividió la imagen digital de manera automática en pequeñas imágenes individuales de 200 x 200 píxeles que se centran en una célula escamosa a alta resolución. Con este conjunto de imágenes digitales recortadas de células escamosas individuales, el patólogo ha clasificado cada célula suelta en cada una de las categorías diagnósticas para obtener una base de datos clasificada y ordenada de cada categoría.

De este modo, hemos generado una base de datos de imágenes de células sueltas individuales, que a modo de fotografías, han sido clasificadas en cada categoría diagnóstica (**Figura 4.3**). Tras esta primera fase nuestra base de datos inicial contenía: 450 imágenes de células benignas, 213 imágenes LSIL, 323

imágenes de ASC-US y 419 imágenes de HSIL. En total, hemos recogido 1405 células reales en la base de datos.

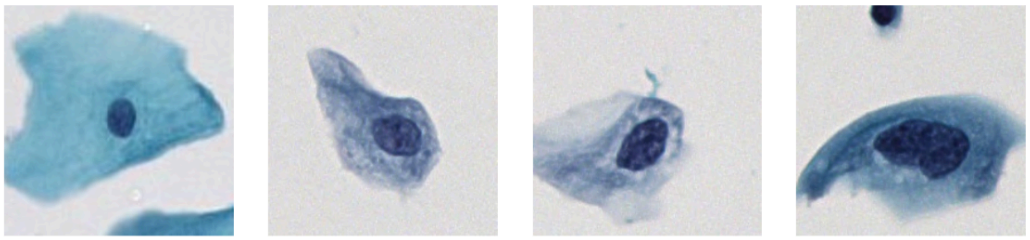


Figura 4.3. Imágenes individuales recortadas del escaneado de todo el portaobjetos. Estas células fueron clasificadas por un patólogo experto en las siguientes categorías (de izquierda a derecha): célula benigna, ASC-US, LSIL y HSIL. Imagen tomada de la tesis doctoral "Clasificación automática de tumores de ovario y citologías cervicovaginales a partir de imágenes ecográficas y microscópicas mediante su análisis con técnicas de aprendizaje automático" del Dr. Martínez Más.

Estas imágenes de células individuales pueden mostrar plegamientos, superposiciones celulares, defectos de tinción o artefactos de procesamiento (**Figura 4.4**). Estos defectos en la citología son muy frecuentes en la práctica diaria y al escanearlos dan lugar a una borrosidad de la imagen, debido a que se produce una superposición de distintos niveles celulares. Esta superposición celular puede solventarse en el visionado manual con el microscopio ya que dispone de una rueda micrométrica que puede modificar el plano de enfoque celular, pudiendo cambiar el plano de enfoque de aquellas células que están amontonadas permitiendo su visionado individual. Sin embargo, en la imagen digital este amontonamiento no se puede evitar, debido a que el escaneado se realiza generalmente en un único foco de ampliación, por lo que las células que estén por delante o detrás de dicho plano, quedarán desenfocadas, ocultas y borrosas.

En algunos trabajos se tiende a realizar un preprocesamiento de las imágenes digitales para tratar de optimizar la calidad de la imagen y reducir los efectos del solapamiento y artefactos anteriormente descritos, ya que facilita la tarea de clasificación cuya exactitud es muy alta (Godinez et al., 2017). Sin embargo, en nuestro trabajo estas imágenes de células individuales no han recibido ningún

preprocesamiento previo; es decir, son directamente escaneadas a alta resolución en condiciones reales. La razón de esta decisión es que al preprocesar la imagen, las células dejan de estar en condiciones reales a las de una citología convencional de una paciente y los resultados obtenidos al entrenar al algoritmo, serían difícilmente aplicables a las citologías de la rutina diaria.

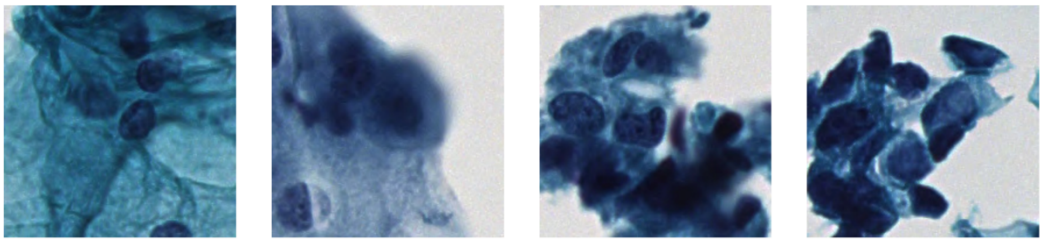


Figura 4.4. Ejemplos en los que la imagen de la célula es de difícil valoración por presentar solapamientos, plegamientos o borrosidad. De izquierda a derecha: célula benigna, ASC-US, LSIL y HSIL. Imagen tomada de la tesis doctoral "Clasificación automática de tumores de ovario y citologías cervicovaginales a partir de imágenes ecográficas y microscópicas mediante su análisis con técnicas de aprendizaje automático" del Dr. Martínez Más.

De este modo, el planteamiento del proyecto es conseguir unos buenos resultados en la clasificación celular, considerando todos aquellos factores que puedan alterar la calidad del visionado de nuestras células, que son tan comunes en la práctica asistencial.

4.1.2. BASE DE DATOS MEJORADA CON CÉLULAS ARTIFICIALES

El DL permite crear modelos de clasificación automática que facilitan la tarea del diagnóstico del cáncer a los patólogos, disminuyendo tiempos de respuesta y optimizando los flujos de trabajo. También sirve como herramienta de apoyo diagnóstico, ya que el diagnóstico del cáncer de cérvix por citología contiene una variabilidad inter e intraobservador no desdeñable.

En nuestro estudio preliminar se señaló que este tipo de modelos de clasificación por imagen presentan la dificultad de obtener una cantidad de células escamosas suficiente para lograr un modelo robusto con una precisión en el diagnóstico comparable o superior a la de un patólogo cualificado. Estas limitaciones de la base de datos no solo son a nivel cuantitativo, debido a la cantidad insuficiente de imágenes celulares, sino también cualitativas, por la complejidad que implica depender del patólogo para el etiquetado manual de cada célula en bases de datos grandes (Martínez-Más et al., 2020).

Además, resulta complejo aumentar de manera equilibrada el número de células reales por categoría diagnóstica, ya que una misma citología rara vez contiene células representativas de todas las categorías y; cuando lo hace, estas se presentan en cantidades desiguales. Por ejemplo, la categoría "benigna" es la más fácil de recolectar para la base de datos debido a que es el diagnóstico más frecuentemente emitido y recoge un gran número de células por portaobjetos, pudiendo encontrar hasta 10.000 células benignas. Sin embargo, diagnósticos como ASC-US son menos prevalentes y representan un pequeño número de células dentro del portaobjetos (<5% de la celularidad total) que se mezclan con células benignas y otras células glandulares e inflamatorias. Como consecuencia, nuestro estudio preliminar partió de una cantidad limitada de células en algunas de las categorías diagnósticas (Martínez-Más et al., 2020).

Para solventar estas dificultades se propone aumentar nuestra base de datos mediante el desarrollo de una base de datos sintética de técnicas artificiales. Estas nuevas células se obtendrán mediante la elaboración de un modelo que combina las redes neuronales antagónicas con las redes profundas convolucionales especializadas en el diagnóstico por imagen. Mediante este nuevo modelo buscamos la creación artificial imágenes de células de similares características a las reales de cada una de las categorías diagnósticas (imitando características citológicas como tamaño celular, color del citoplasma o forma y tamaño del núcleo) para así enriquecer la muestra y perfeccionar el algoritmo del DL y los resultados obtenidos en nuestro estudio preliminar (Martínez-Más et al., 2020).

4.2. CREACIÓN DE CÉLULAS ARTIFICIALES

El desarrollo de células artificiales se ha llevado a cabo mediante el uso de una red antagónica o red generativa adversaria. Las redes antagónicas generativas (GAN, por sus siglas en inglés) son técnicas de ML no supervisado que se emplean para generar datos o imágenes artificiales que sean muy similares a datos reales con los que la red fue entrenada. Una red GAN está compuesta a su vez por dos redes diferenciadas que compiten ente sí con un objetivo común:

1. Un generador de imágenes que crea datos artificiales que parezcan reales. En nuestro proyecto, se tratan de imágenes de células escamosas cervicales artificiales nunca vistas, que sean similares a las reales.

2. Un discriminador que trata de diferenciar que datos o imágenes son reales y cuales artificiales. Para ello se entrena previamente con fotos reales para conseguir que sea capaz de detectar las artificiales.

Ambas redes compiten entre sí, por eso se conocen como redes antagónicas. La red que actúa como generador trata de confundir a la red discriminadora haciéndose pasar por una imagen real; mientras que la red discriminadora trata de detectar aquellos datos que son artificiales generados por ordenador. Cuanto más se entrena la red, tanto el generador como el discriminador mejora su tarea por lo que se obtienen resultados cada vez más realistas.

Para diseñar una GAN, primero hay que entrenar a la red con imágenes reales, en nuestro caso con nuestra base de datos de células reales de cada una de las categorías diagnósticas del Sistema Bethesda. Esta fase de entrenamiento afecta a ambas redes, para que después el generador cree células artificiales y se mejoren los resultados tras la acción del discriminador.

Para las redes antagónicas de este modelo se ha empleado una red StyleGAN2, que se trata de una mejora de la red original StyleGAN, diseñada por NVIDIA, que genera imágenes artificiales muy realistas a partir de redes GAN (**Figura 4.5**). La red StyleGAN2 destaca por su capacidad de generar imágenes realistas muy detalladas de rostros humanos que no pertenecen a personas reales, paisajes o diseño de objetos. En nuestro proyecto hemos empleado una red GAN de este tipo, resultando distinta para cada diagnóstico, sin proporcionarle información sobre el resto de categorías. Es decir, hemos empleado una red StyleGAN2 distinta para cada categoría diagnóstica, debido a que nos interesa que cada célula artificial generada cumpla los criterios y características necesarias que justifiquen su clasificación en la categoría a la que pertenece, siendo en ocasiones las diferencias entre imágenes de distintas categorías muy sutiles.

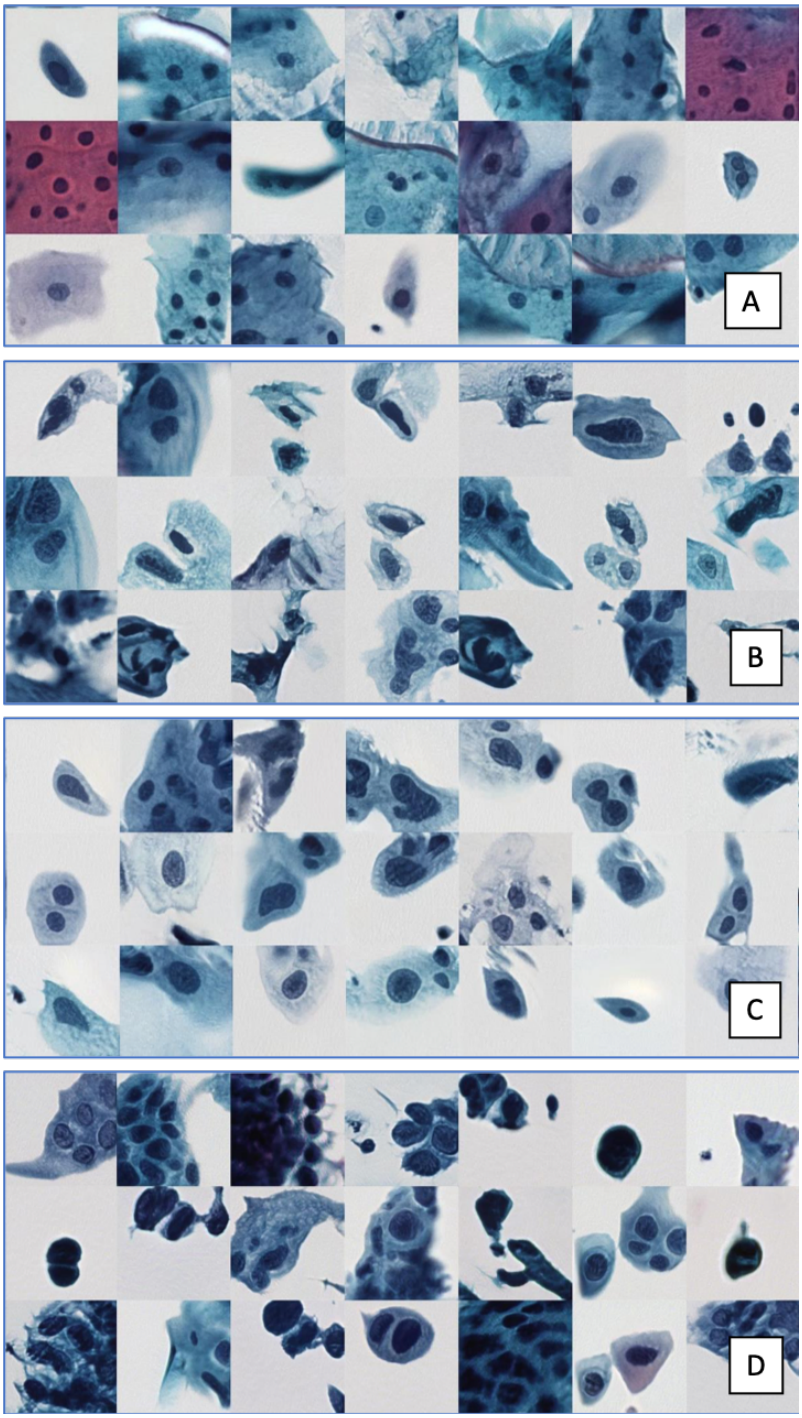


Figura 4.5. Conjunto de células artificiales generadas por la red StyleGAN2 que incluyen tanto las células correctamente generadas como las mal generadas: (A) benignas, (B) LSIL, (C) HSIL y (D) ASC-US. Elaboración propia.

De este modo hemos empleado 4 redes StyleGAN2 que hemos entrenado con imágenes de células reales; empleando para el entrenamiento de cada red células reales únicamente de la categoría diagnóstica para la que han sido generadas (benigna, ASC-US, LSIL y HSIL). A través de las redes antagónicas se espera generar células artificiales con características citológicas similares a las células reales con las que fueron entrenadas, que permitan clasificar la célula artificial en la categoría diagnóstica para la que fue creada por un patólogo experto. De esta forma al obtener células artificiales indistinguibles de las reales por un patólogo cualificado y con características citológicas fieles al Sistema Bethesda, podremos crear una base de datos más amplia que permita obtener un modelo de diagnóstico celular más robusto.

4.3. ELABORACIÓN DE UN MODELO HÍBRIDO PARA LA CLASIFICACIÓN DE IMÁGENES

La aplicabilidad del sistema de aprendizaje profundo convolucional (DCNN) ha sido probada en el campo de la medicina obteniendo buenos resultados. Se han empleado redes neuronales para evaluar el grado de severidad de neumonías en pacientes con COVID-19, mediante el análisis de las radiografías de tórax, obteniéndose hasta un 93,1% de efectividad (Abbas et al., 2021). También se ha desarrollado un modelo DCNN fiable en la clasificación de imágenes de distintos subtipos del cáncer de mama mediante la evaluación de tejidos de pacientes teñidos con hematoxilina-eosina (Motlagh et al., 2018). En el campo de la citología también se ha tratado de aplicar redes para generar células artificiales; sin embargo, son imágenes de células artificiales poco realistas ya que los autores realizaron una preparación previa de la muestra eliminando los artefactos de fondo, microorganismos o células inflamatorias acompañantes que no son objeto de interés (Godinez et al., 2017). En nuestro modelo tratamos de diseñar nuevas células lo más realistas posibles, incluyendo artefactos, células inflamatorias

acompañantes, problemas en la tinción, superposición de células o plegamientos del núcleo o citoplasma; acercándolos lo máximo a la realidad.

Tras la creación de una nueva base de datos de células artificiales generadas por la red antagónica (GAN), la hemos fusionado a una red DCNN, obteniendo una nueva red híbrida fusionada, llamada DCGAN, del inglés *Deep Convolutional Geberative Adversarial Network*. Se espera que esta red sea capaz de etiquetar estas nuevas células artificiales creadas con una precisión similar a la obtenida por patólogos experimentados en el etiquetado de células reales en la primera parte del estudio.

Para la red DCNN se empleó el modelo Xception, una arquitectura avanzada preentrenada con la base de datos ImageNet, la cual contiene imágenes ajenas al ámbito médico. Se optimizaron sus parámetros y se aplicó aprendizaje por transferencia, reutilizando los pesos obtenidos a partir de ImageNet para realizar un nuevo preentrenamiento con imágenes de células cervicales. De este modo, tras esta fase la red ha adquirido la capacidad de clasificar células en distintas categorías diagnósticas, tras la que le hemos introducido las células artificiales para testear el modelo y ver que diagnósticos emite en su capa de salida. Las células mal clasificadas en una categoría errónea pueden ser estudiadas para ajustar los parámetros de la red StyleGAN2 para que la red antagónica pueda afinar los parámetros y sea capaz de discernir entre aspectos más finos que pueden ser similares entre las categorías para así disminuir los errores.

Para comprobar la capacidad y veracidad del modelo híbrido de crear y clasificar células correctamente, hemos subclasificado nuestra base de datos de células artificiales en las distintas categorías diagnósticas del Sistema Bethesda mediante su predicción automatizada por el modelo híbrido, para verificar si son clasificadas en la categoría diagnóstica para la que fueron creadas.

Al comparar la clasificación humana con la clasificación automática de la red DCNN, podemos no solo evaluar la capacidad de la red de clasificar las células en su categoría correcta, sino también la capacidad de la StyleGAN2 de generar células

artificiales y ver si son superponibles a las células reales de cada categoría. Al final de esta fase del estudio solo hemos conservado las células correctamente generadas y clasificadas, descartando todas las demás (**Figura 4.6**).

De este modo, el modelo híbrido (DCGAN) está compuesto por (**Figura 4.7**):

- Una red antagónica (GAN) que generará nuevas imágenes de las células para cada categoría diagnóstica, mediante una red StyleGAN2.
- Una red neuronal DCNN que verificará si la nueva imagen diseñada corresponde con la categoría diagnóstica para la que ha sido creada, mediante un modelo de red Xception.

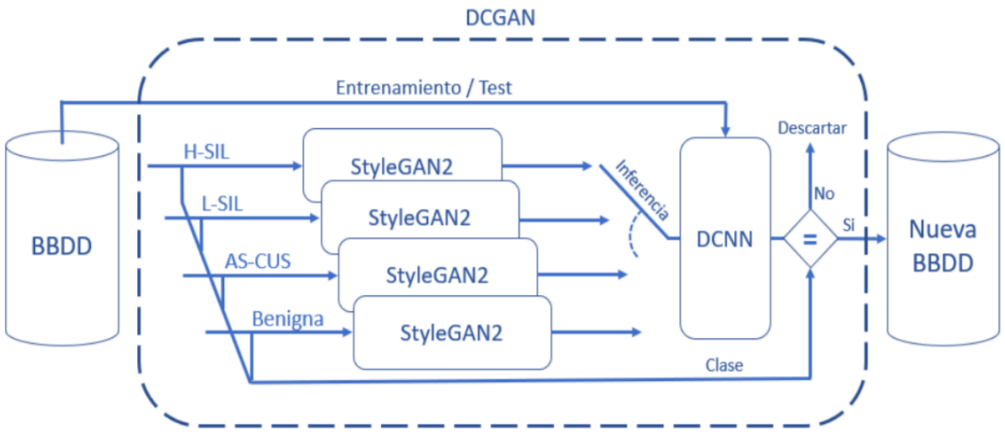


Figura 4.6. Modelo híbrido propuesto en el que se combina una red GAN y una DCNN para obtener una nueva base de datos amplia y mejorada. Elaboración propia.

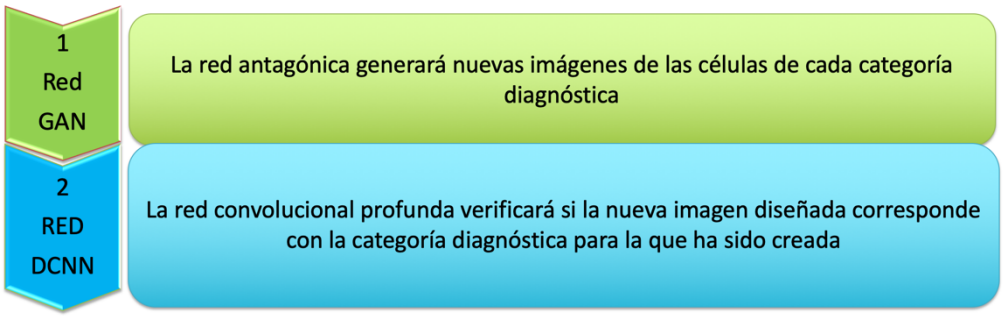


Figura 4.7. Estructura del modelo híbrido DCGAN. Elaboración propia.

4.4. LA SEGMENTACIÓN COMO AYUDA AL DIAGNÓSTICO POR IMAGEN

Los modelos de IA interpretables son aquellos cuyo funcionamiento puede ser entendido por una persona gracias a que las decisiones del modelo son explicadas, facilitando la comprensión de cómo funcionan internamente. Un ejemplo de este tipo de modelos es el árbol de decisión ya que se puede seguir el camino por el que el modelo toma la decisión. Es decir, son características intrínsecas del modelo que permiten saber qué camino ha llevado para elaborar su predicción final en la capa de salida, de modo que un observador externo puede entender los pasos que ha seguido el modelo desde su entrada de datos hasta su toma de decisión. Por otro lado, los modelos de IA explicable son aquellos que se centran en justificar su decisión final, independientemente del camino que ha llevado a la red a dicha decisión. Ambos métodos son críticos para mejorar la confianza en la IA, ya que le aportan transparencia a la red para mejorar su comprensión por los profesionales expertos.

La segmentación de imágenes es una rama de la IA que divide una imagen completa en distintas regiones. El uso de este método sobre imágenes médicas presenta una gran utilidad, debido a que la segmentación automática hace más fluido el procesamiento de datos y ofrece visualizaciones específicas de áreas de la imagen, resultando de gran utilidad a los médicos permitiendo realizar mediciones específicas de área permitiendo; por ejemplo, monitorear sobre una imagen la progresión del cáncer (Shao et al., 2023).

Aplicado a nuestro trabajo, proponemos el uso de la segmentación con la elaboración de máscaras celulares como herramienta de interpretabilidad que aporte claridad en la toma de decisiones de la red, mostrándonos las zonas de interés de la célula que son de mayor importancia para emitir un diagnóstico.

Mediante la elaboración de máscaras celulares como método de segmentación podemos entrenar al algoritmo de una manera más dirigida y optimizada en vistas a una nueva fase de clasificación celular. A través de la segmentación tratamos de enriquecer la imagen de entrada a la red con la elaboración de máscaras celulares que marcan las áreas más importantes de la célula que son fundamentales para el etiquetado (el núcleo y el citoplasma), que le sirvan a la red para mejorar su precisión en el diagnóstico celular. De este modo esperamos una mejora de la precisión del modelo en el diagnóstico del cáncer de cérvix y una mejora de la confianza de los profesionales sanitarios este tipo de modelos.

4.4.1. PREPARACIÓN DE NUESTRA BASE DE DATOS PARA SEGMENTACIÓN

Hemos empleado células reales obtenidas de citologías escaneadas del Complejo Hospitalario Universitario de Cartagena, seleccionadas para nuestro estudio preliminar, de 1405 células reales (Martínez-Más et al., 2020) distribuidas entre las distintas categorías diagnósticas del Sistema Bethesda tras su clasificación por un patólogo en: 450 imágenes de células benignas, 213 imágenes LSIL, 323 imágenes de ASC-US y 419 imágenes de HSIL. Se aplicaron algunas técnicas de aumento de datos mediante transformaciones sobre la imagen para tratar de equilibrar la muestra al grupo de mayor cantidad de células, como: zoom de hasta 15%, rotaciones de 10 grados, desplazamientos (horizontales y verticales) de hasta el 20%, volteos de la imagen (horizontales o verticales) y recortes o distorsiones de hasta el 15%. Para corregir cualquier espacio vacío creado por estas transformaciones, el píxel más cercano rellenará el área mediante estiramiento.

Del mismo modo a otras fases del proyecto trabajamos con imágenes no preprocesadas para acercarnos lo máximo posible a las condiciones diagnósticas reales, aunque artefactos como solapamiento y borrosidad en la imagen dificulten y encarezcan la segmentación de la imagen y represente un reto de clasificación celular mayor para el algoritmo.

Para nuestra prueba de segmentación con elaboración de máscaras celulares, patólogos expertos han seleccionado 30 imágenes para cada categoría diagnóstica. Con el grupo de células sobrante, el 80% se ha empleado para entrenamiento y el 20% restante para validación.

4.4.2. DISEÑO DE NUESTRO MODELO DE SEGMENTACIÓN Y CLASIFICACIÓN CELULAR

Tras la recogida de la base de datos, planteamos un modelo que combina segmentación (mediante la elaboración de máscaras de imagen para cada célula a través de un método semiautomático) y clasificación celular. Para ello hemos seguido las siguientes fases.

1) La segmentación manual de una muestra representativa con la elaboración de máscaras celulares por un patólogo experto.
2) La segmentación automática para las imágenes restantes con la elaboración automática de máscaras celulares por la red.
3) El sometimiento a la red a una nueva etapa de clasificación para estudiar si mejora sus resultados tras la etapa de segmentación.

4.4.2.1. ***Fase 1***

En primer lugar, hemos seleccionado 20 imágenes de células individuales de cada categoría diagnóstica (con un total de 80) y un patólogo ha delimitado de forma manual el núcleo y citoplasma mediante un sistema colorimétrico que los diferencia, generando una máscara celular (**Figura 4.8**). Los núcleos se han marcado de rojo y los citoplasmas de azul, despreciando el fondo de la imagen (regiones externas de la célula) (**Figura 5.5**).

Este sistema tiene la ventaja de delimitar en la célula las zonas relevantes para el diagnóstico y de mejorar la calidad de la imagen celular, despreciando áreas borrosas, de plegamiento o de solapamiento celular. La segmentación de la imagen actúa con la misma lógica que emplea el patólogo, ya que para emitir un

diagnóstico se centra en las características del núcleo y citoplasma celular, intentando identificar criterios citológicos que estén recogidos en el Sistema Bethesda y le permitan clasificarla, obviando otros elementos de la imagen que no determinan el diagnóstico, como celularidad inflamatoria de fondo, hematíes, microorganismos o artefactos.

Estas características que orientan a que la célula tiene riesgo potencial de desarrollar cáncer quedan reseñadas tras la elaboración de las máscaras celulares empleando el sistema de colores. Por ejemplo, según el Sistema Bethesda; de forma general, el aumento del tamaño del núcleo respecto al citoplasma (conocido como el aumento de la relación núcleo-citoplasma) orienta a que la célula puede ser displásica. Otro hallazgo, como la presencia de núcleos de contornos irregulares en lugar de redondeados también orienta a displasia.

4.4.2.2. ***Fase 2***

Posteriormente a través del entrenamiento del algoritmo con estas imágenes enriquecidas con una máscara manual, se trata de obtener la segmentación automática celular.

Para la generación automática de las máscaras de las imágenes restantes, se ha empleado una DCNN preentrenada en la clasificación de imágenes y elaboración de máscaras para evitar largos periodos de entrenamiento y asegurar una buena comprensión del modelo. El modelo que mejor se adapta a la segmentación y generación automática de máscaras es una red totalmente convolucional (FCN).

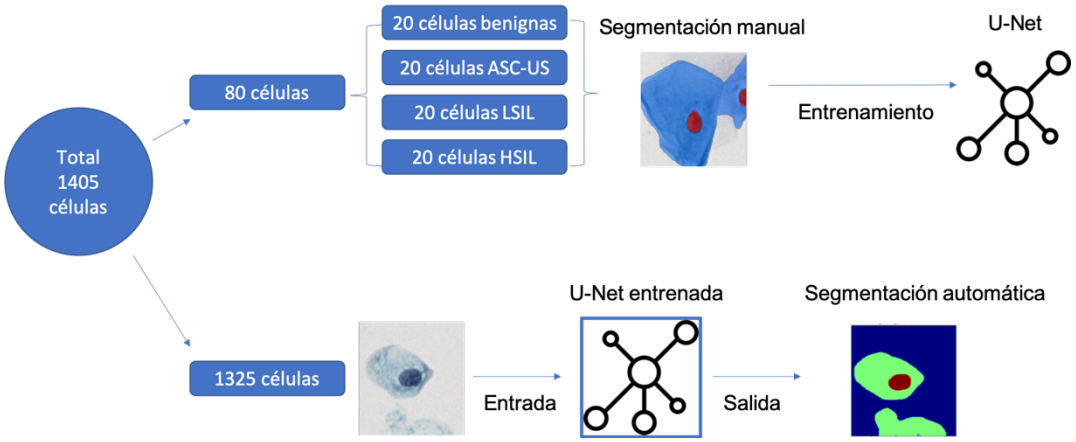


Figura 4.8. Fases 1 y 2 del diseño del modelo de segmentación automática. En primer lugar se prepara una base de datos de células segmentadas manualmente para entrenar a la red U-Net. Posteriormente se utilizan nuevas imágenes no testeadas previamente como imagen de entrada a la espera de obtener una célula con una máscara generada automáticamente en la capa de salida. Elaboración propia.

Las redes totalmente convolucionales o FCN; al igual que las redes CNN, tienen una capa de entrada, varias capas intermedias y una de salida. Sin embargo, y al contrario que las redes CNN, la FCN emplea *MaxPooling* que reduce el tamaño de la imagen tras extraer previamente las características más importantes de la imagen original. El *MaxPooling* busca reducir el tamaño de la imagen para disminuir la cantidad de cálculos y conseguir que el modelo sea más eficiente, pero conservando las características más importantes de la imagen, logrando que dichas características sean las que pasan a las siguientes capas.

Las principales ventajas de este sistema es que le aporta robustez al modelo ante ruidos o variaciones de la imagen y reduce el riesgo de sobreajuste. Estos modelos de segmentación presentan el inconveniente de que a medida que se extraen las características, el tamaño de la imagen de salida se reduce, generándose una pérdida en la definición de la misma con la consecuente elaboración de mapas de características menos precisas a mayores profundidades. Para solventar esta desventaja, hemos empleado un decodificador al final del proceso para recuperar la imagen original.

Entre los modelos disponibles desarrollados a partir de FCN, hemos seleccionado la *U-Net*. Se trata de una red diseñada para la segmentación de imágenes biomédicas, que fue creada inicialmente para identificar tumores de pulmón y cerebrales, empleado por Ronneberger y cols. (Ronneberger et al., 2015). La *U-Net* contiene dos etapas: (1) etapa de contracción o codificación en la que la imagen se va reduciendo progresivamente para extraer sus características profundas y (2) una fase de expansión o decodificación donde se devuelve la imagen a su tamaño original. En la etapa de contracción se aplican convoluciones que buscan patrones en la imagen, o *pooling* que a la vez que reducen el tamaño de la imagen conservan la información más relevante de la misma. Hay estudios que han aplicado la U-Net al diagnóstico de citologías cervicovaginales, observando una discreta mejora de las métricas (un 0,41% de mejora en precisión y 1,3% en valor F1) respecto a métodos tradicionales (Albzour & Lam, 2025).

Este modelo preentrenado para el diagnóstico de dichas neoplasias se ha reentrenado con el 5% de nuestras imágenes celulares segmentadas manualmente. Una vez la U-Net ha aprendido a segmentar células, le hemos dado una nueva entrada con imágenes originales a la espera de obtener células con máscaras generadas automáticamente en la capa de salida (**Figura 4.8**).

4.4.2.3. ***Fase 3***

Por último, una vez que la red U-Net adquiere la capacidad de generar máscaras automáticas tras su entrenamiento con un 5% de las células segmentadas con máscaras manuales, la fusionamos con una red Xception a través de un *clasificador* para que la red fusionada emita un diagnóstico final con una de las categorías del Sistema Bethesda (**Figura 4.9**). Testeamos varios *clasificadores* para fusionar las redes que se discuten en el siguiente apartado.

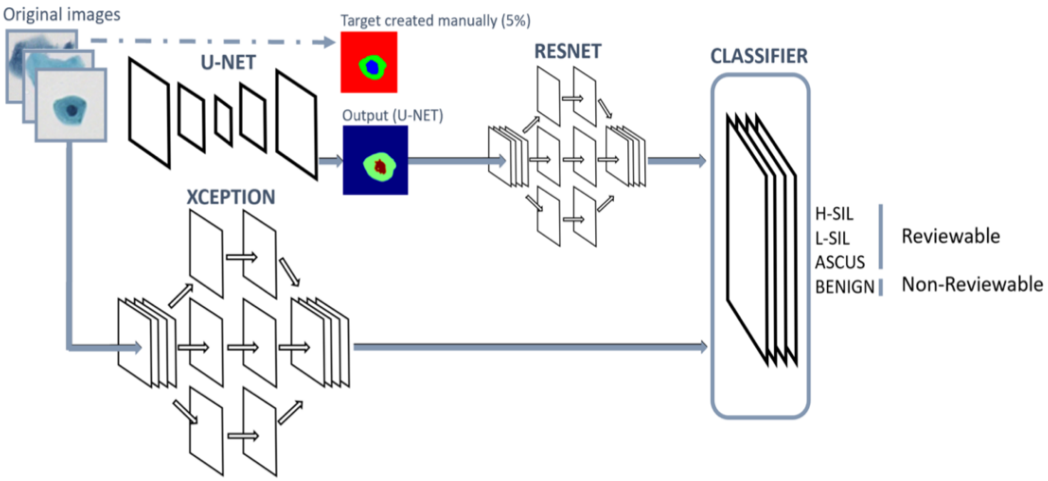


Figura 4.9. Fase 3 del modelo propuesto. Una vez que la red aprende a segmentar automáticamente, esta se fusiona con una red Xception mediante un clasificador. Dependiendo del modelo de clasificador, proponemos distintos modelos. Elaboración propia.

4.4.3. EL EMPLEO DE DISTINTOS CLASIFICADORES EN LOS MODELOS DE SEGMENTACIÓN Y CLASIFICACIÓN CELULAR PROPUESTOS

Diseñamos dos modelos de segmentación y clasificación celular que integran la segmentación y extracción de características, procesando imágenes y máscaras de segmentación en paralelo. Ambos modelos se diferencian entre sí por el clasificador empleado para fusionar la red U-Net de segmentación con la red Xception que emitirá un diagnóstico celular. En el capítulo correspondiente compararemos los resultados generados por estas redes propuestas con los obtenidos por un modelo clásico de clasificación de imágenes que no integra la segmentación, para evaluar qué modelo tiene mejor rendimiento.

4.4.3.1. ***Modelo clásico***

El modelo clásico empleado está basado en una red Xception preentrenada con imágenes de ImageNet, que contiene 40 capas congeladas que retienen las características generales esenciales de la imagen, para que en capas posteriores se especialice en matices concretos de la imagen celular. Este método de *transfer learning* que emplea una red preentrenada, tiene la ventaja de que reduce el tiempo de entrenamiento y aprovecha el conocimiento previo de la red. Para transferir los conocimientos previos de la red se congelan sus capas; es decir, se bloquean para que no actualicen los pesos ni aprendan durante el entrenamiento con nuevos datos, de modo que se quedan estancos con lo aprendido de su entrenamiento anterior. En nuestro estudio hemos congelado las primeras 40 capas, las cuales evalúan las características generales de la imagen como textura, bordes o formas; permitiendo a su vez que las capas posteriores se especialicen en matices o tareas específicas de la imagen, siendo capaces de distinguir detalles más finos para adaptarlos a su nueva tarea de clasificación.

4.4.3.2. ***Modelo 1***

El primer modelo que proponemos integra la red de segmentación U-Net con Xception, permitiendo el procesamiento simultáneo de segmentación (U-Net) y extracción de características (Xception). Por un lado, la red Xcpetion realiza una predicción del diagnóstico sobre células originales al igual que en otras fases del estudio, y por otra la U-Net segmenta las células para posteriormente emplearlas como imagen de entrada en otra clasificación en paralelo.

A diferencia del modelo clásico, éste modela imágenes y máscaras de segmentación en paralelo. El componente U-Net segmenta la imagen y trabaja con las características tanto de la imagen segmentada como de la imagen original, lo que aumenta la capacidad del modelo para capturar patrones celulares complejos y mejora el rendimiento tanto en tareas de segmentación como de clasificación.

Posteriormente, la salida de U-Net se conecta con una red ResNet o *Residual Network* que aprovecha las propiedades de preservación de gradiente para reducir la pérdida de información, importante para capturar detalles de anomalías celulares cruciales para aplicaciones que requieren una segmentación de alta precisión. La ResNet usa conexiones residuales o atajos que saltan capas en redes muy profundas y suma la entrada original con la salida transformada obtenida, de manera que la información original se conserva a lo largo de la red. La U-Net recibe la célula segmentada emitida previamente por la U-Net como imagen de entrada y emite un diagnóstico final en su capa de salida.

Por último, un clasificador recibe la salida en paralelo del diagnóstico de la red Xception emitido sobre la imagen celular real y el diagnóstico de la ResNet realizado sobre la célula segmentada y emite un diagnóstico celular combinado de ambas predicciones. En este modelo el clasificador utilizado emplea una capa aplanada y una capa densa (**Figura 4.10**). La capa aplanada se encarga de convertir toda la información multidimensional obtenida de la imagen en un vector unidimensional largo que se introduce en la capa densa que es capaz de interpretar dicha información y emitir un diagnóstico final. La capa densa se caracteriza porque cada neurona esta conectada con todas las neuronas de la capa anterior.

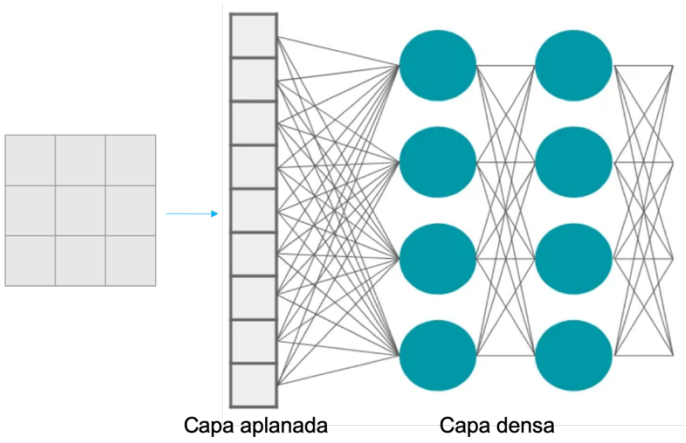


Figura 4.10. La capa aplanada convierte la información en un vector plano que presenta las características necesarias para su lectura por la capa densa que tiene todas las redes conectadas entre sí. Elaboración propia.

Este clasificador tiene el inconveniente de que su eficiencia computacional no está optimizada respecto a otros clasificadores, debido a la complejidad de su arquitectura. Al manejar más parámetros en un vector largo, presenta mayor riesgo de sobreajuste y requiere tiempos más largos de entrenamiento.

4.4.3.3. ***Modelo 2***

El segundo modelo que proponemos contiene la misma estructura que el modelo anterior al cual perfecciona incorporando una capa GAP (*Global Average Pooling* o Agrupación Media Global) como mecanismo de clasificación (**Figura 4.11**). La capa GAP realiza un único valor promedio de los mapas de características generados por la última capa convolucional justo antes de la clasificación final, permitiendo relacionar el resultado únicamente con una clase.

Al trabajar con valores promedios, disminuye la dimensionalidad del modelo y mejora su eficiencia computacional. Este sistema reduce la cantidad de parámetros, preservando los datos espaciales sin aplanar los mapas de características, permitiendo el procesamiento eficiente de relaciones espaciales complejas, minimizando el riesgo de sobreajuste, acelerando el proceso de entrenamiento y aportando robustez al modelo.

La capa GAP reduce la cantidad de datos manteniendo la información importante justo antes de la clasificación final, mientras presenta la misma capacidad de procesar la segmentación de U-Net y de extraer las características de la imagen obtenidas mediante Xception que el modelo anterior.

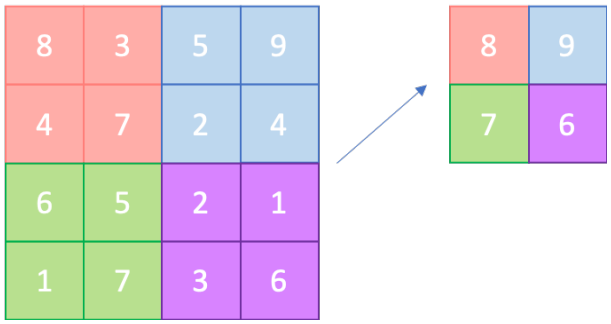


Figura 4.11. Capa GAP como clasificador. Realiza un valor promedio de los mapas de características generados reduciendo significativamente la cantidad de datos. Elaboración propia.

De este modo planteamos estudiar y comparar los 3 modelos, presentando el modelo clásico como un modelo robusto y eficiente que emplea una red Xception como método de clasificación clásica. Un modelo 1 que produce una segmentación y clasificación paralelas al incorporar una U-Net y ResNet mejorando al modelo clásico. Finalmente, un modelo 2 que optimiza al anterior por incorporar una capa GAP que reduce la dimensionalidad y mejora la eficiencia computacional, haciéndolo ideal para aplicaciones de alta precisión y a gran escala.

4.5. LA EXPLICABILIDAD EN LOS MODELOS DE DIAGNÓSTICO POR IMAGEN

Las técnicas de DL, en particular las CNN, han demostrado una capacidad diagnóstica notable. No obstante, a pesar de su alta precisión en el etiquetado celular, presentan una limitación importante: su escasa explicabilidad. Estos modelos se comportan como cajas negras, que no aportan información sobre los motivos que justifican su decisión final. Esta problemática tiene un gran peso en el

campo diagnóstico de la medicina, ya que a pesar de que se tratan de sistemas con gran capacidad computacional, la ausencia de explicabilidad dificulta mucho su implantación en la actividad asistencial debido a la falta de confianza del personal sanitario ante estos modelos.

La inteligencia artificial explicable (XIA), es un tipo de IA que desarrolla sus razonamientos o razones que permitan entender por qué toma las decisiones. Este tipo de métodos son de gran interés en el campo de la salud, ya que al proporcionar las razones de la toma de decisiones del modelo, se mejora la confianza del personal sanitario en el mismo. Del mismo modo, se puede emplear esa información tanto en el aprendizaje del modelo como de los facultativos expertos.

Se ha experimentado con varios tipos de XIA en el diagnóstico de imágenes médicas mediante resonancias magnéticas, tomografías computarizas o radiografías. Los mapas de calor, se plantean como una manera sencilla de marcar sobre la imagen las zonas de interés que sientan las bases de la predicción del modelo; como los Grad-CAM (mapa de activación de clases por ponderado por gradiente) o Saliency Maps. Existen otros modelos de explicación local (LIME o SHAP) que mediante pequeñas explicaciones locales; como el desarrollo de las características de la imagen o la cuantificación de cuánto ha contribuido cada sección de la imagen en el modelo, tratan de explicar el resultado de la predicción. Otros tipos de modelos de XIA son las redes de decisión o las redes explicables con representación de casos, que a través de una imagen de referencia es capaz de detectar un caso similar en una imagen nueva.

Por lo tanto, para contrarrestar esta desventaja de la falta de transparencia de las decisiones del modelo, planteamos el uso de XIA en el diagnóstico de imágenes celulares y comparamos su precisión con la obtenida por los modelos de redes CNN convencionales no enriquecidas con mecanismos de explicabilidad.

4.5.1. APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL EXPLICABLE EN CITOLOGÍAS CERVICOVAGINALES

En esta nueva fase del proyecto se propone un nueva arquitectura híbrida que incorpore un método de explicabilidad que pueda explicar los resultados obtenidos por el modelo en el diagnóstico de células. Para ello tratamos e evaluar cuál es el modelo que mejor se adapta y qué método de explicabilidad puede aportarnos mejores resultados. Además, trataremos de usar la información obtenida a través del modelo explicable para mejorar los resultados de la red en una nueva clasificación de imágenes.

Para elaborar el modelo híbrido, hemos seleccionado una red especializada en el reconocimiento de imágenes tipo Xception, preentrenada con imágenes de ImageNet en las que aprende las características generales de una amplia variedad de imágenes distintas a las del campo médico. Esta red analiza paso a paso las imágenes organizándolas en secciones según sus características como forma, tamaño o colores. Mediante la técnica *transfer learning* hemos aprovechado la arquitectura y conocimientos previos del modelo para su nueva tarea y a través del *fine tuning;* mediante ajustes adicionales de hiperparámetros y agregando varias capas de salida, hemos adaptado el modelo al problema específico de clasificación celular. De este modo el algoritmo aprovecha sus conocimientos previos de clasificación de imágenes y los aplica a la clasificación de imágenes de células.

Posteriormente, tras entrenar a la red Xception con imágenes de células, hemos combinado esta red con una técnica de explicabilidad para construir un modelo híbrido robusto. Se han testeado diversas técnicas de explicabilidad que podrían tener cabida en nuestro estudio:

- Sensibilidad a la oclusión: consiste en ocluir mediante un cuadrado gris o "parche" distintas partes de la imagen. Tras la oclusión se clasifica la imagen modificada y se calculan las probabilidades de pertenencia a la clase. Si ante un determinado parche disminuyen notablemente estas

probabilidades se concluye que dicha área influye significativamente en la clasificación. Tras esto, se genera un mapa de calor que resalta las posiciones de los parches que tienen un mayor impacto en la clasificación. Este sistema se ha empleado en algunos campos de la rama sanitaria, como en el estudio del envejecimiento celular mediante el análisis de imágenes cerebrales, para tratar de estudiar qué áreas del cerebro actúan como biomarcador del envejecimiento (L. He et al., 2025).

- BagNet: clasifica las imágenes utilizando la información generada por pequeños parches locales sin considerar el aspecto global de la imagen. Una vez obtenido los resultados independientes, se combinan para obtener una predicción final. Finalmente se elabora un mapa de calor que pone en evidencia la importancia de cada parche. La BagNet es un modelo IA explicable que se ha empleado para el diagnóstico de la membrana epirretiniana (capa de tejido fibroso que se forma en una de las capas de la retina que produce distorsión de la visión) en imágenes de tomografía de coherencia óptica, demostrándose que las explicaciones visuales que el modelo BagNet ofrece para sus decisiones, se ajustan a la perfección a la experiencia clínica, sirviendo como guía para las decisiones clínicas en la práctica asistencial (Mensah et al., 2025).

- Mapeo por activación de clases ponderado por gradiente o Grad-CAM: se genera un mapa de resalta en forma de gradiente el peso obtenido de cada región de la imagen. Hay publicados varios estudios sobre el empleo del mapa de calor a través del Grad-CAM en la clasificación de imágenes de patología ovárica (Jung et al., 2022), de metástasis de ganglios linfáticos (Suara et al., 2024) y del análisis de imágenes histológicas para el diagnóstico de distintos tipos de cáncer de pulmón (Mothkur et al., 2025), con buenos resultados.

- LIME: se basa en alteraciones en la imagen original para tratar de comprender como cambian las predicciones del modelo, proporcionando información sobre la contribución de cada característica a la clasificación.

> Se ha integrado su uso con la red CNN-TumorNet que se encarga de clasificar imágenes de resonancia magnéticas cerebrales en categorías tumorales y no tumorales, aportándole transparencia al modelo, facilitando su implantación en entornos clínicos y mejorando la atención del paciente neurooncológico (Rasool et al., 2025).

En nuestro proyecto nos hemos decantado por este sistema de Grad-CAM con el que mediante un sistema colorimétrico hemos tratado de detectar las zonas de la célula en las que el modelo de DL se ha fijado a la hora de realizar su clasificación. Este sistema de mapeo que emplea un mapa de calor activa en primer lugar las zonas más relevantes de la imagen para a continuación establecer los gradientes, que se resaltan con un sistema colorimétrico. Los gradientes señalan cómo cambia la predicción del modelo si cambiamos ligeramente las características de esa área de la imagen. De esta forma se pone de manifiesto mediante un sistema visual qué tan importante es cada parte de la imagen para llegar a la predicción final. Por ejemplo, si trabajáramos con imágenes de perros, la activación se centraría únicamente en el perro y los gradientes marcarían que parte de la imagen es más importante para la clasificación, como podrían ser las orejas o las patas.

Paso a paso, la explicabilidad del modelo Grad-CAM funciona de la siguiente manera:

1. Activaciones de la red. La red analiza la imagen de la célula cervical, estableciendo una predicción; por ejemplo, "benigna" (no revisable). Al realizar la predicción, la red reacciona con mayor intensidad ante determinadas características de la imagen original, que son las áreas de activación de la red.

2. Cálculo de los gradientes. Continúa analizando, cómo son de sensibles las predicciones del diagnóstico si realizamos cambios en las áreas de activación de la imagen. Tratamos de poner de manifiesto

cómo de importante es cada área de la imagen respecto a las demás para la toma de decisión del algoritmo.

3. Ponderación de las activaciones. Una vez que sabemos como son de sensibles las áreas al cambio en los resultados de la predicción (analizado en el cálculo de los gradientes), les damos un peso según su importancia para la red dándole un color que resultará en un mapa colorimétrico o mapa de calor cuya fusión con la imagen original vendrá regulada por el parámetro α.

El parámetro α, hace referencia al factor de ponderación o fusión que trata de equilibrar la contribución en la predicción final de las activaciones y los gradientes. Nos permite regular cuánto influyen las activaciones y los gradientes en el mapa final, de modo que si el valor de α es alto la red pondrá más atención a los gradientes y si es bajo a las activaciones. De manera práctica, mediante el parámetro α controlamos cuánto de la activación y cuánto del gradiente se combinan en el mapa colorimétrico; es decir, el grado de fusión entre la célula real y el mapa de calor. El parámetro α es importante porque los mapas se pueden beneficiar de un ajuste fino mejorando los resultados según a qué se le de más peso, si a la activación o al gradiente.

De este modo mediante el diseño de este modelo híbrido esperamos; no solo entender las razones por las que el algoritmo emite un diagnóstico celular, sino también pretendemos aprovechar este modelo explicable para mejorar la precisión del algoritmo.

4.5.2. DISEÑO DE UN MODELO HÍBRIDO EXPLICABLE MEDIANTE EL USO DE UNA RED GRAD-CAM CON MAPAS DE CALOR

Planteamos elaborar un modelo híbrido explicable que emplea una red Grad-CAM con mapas de calor, con los siguientes objetivos:

1. Lograr que el modelo consiga extraer un mapa colorimétrico o mapa de calor de la imagen original aportándole explicabilidad al modelo, en la que se vea reflejado de manera gráfica en la capa de salida de la red dónde se ha fijado el algoritmo para tomar su decisión diagnóstica.

2. Estudiar qué modificaciones se producen en las predicciones si fusionamos la imagen de salida del mapa de calor con la célula original y utilizamos esta nueva imagen enriquecida como entrada a un segundo modelo. Tras testear este grado de fusión condicionado por el parámetro α, tratamos de estudiar si se mejora la precisión del modelo en el etiquetado de imágenes al emplear estas imágenes fusionadas enriquecidas.

De este modo, nuestro modelo tiene una arquitectura distribuida en dos etapas **(Figura 4.12)**:

- Etapa de explicabilidad. Esta primera etapa contiene un modelo CNN tipo Xception previamente preentrenado para hacer predicciones de imagen, que tras volver a entrenarlo con imágenes de células, es capaz de extraer el mapa de calor usando el mapeo por activación de clases ponderado por gradiente (Grad-CAM). De esta forma obtenemos una imagen de la célula marcada con un mapa colorimétrico o de calor que vamos a superponer con la imagen de la célula original en función a un porcentaje de ponderación determinado por el parámetro α.

- Etapa de clasificación. Esta imagen de la célula fusionada con el mapa de calor generada en la salida, pasa a formar parte de la entrada de otra nueva red CNN tipo Xception en esta segunda fase, donde se clasifica la célula en una de las cuatro categorías diagnósticas del Sistema Bethesda. Estas clasificaciones iniciales convergen en una clasificación final simplificada de tipo binaria: "revisable" y "no revisable". La clasificación binaria constituirá el resultado final del modelo.

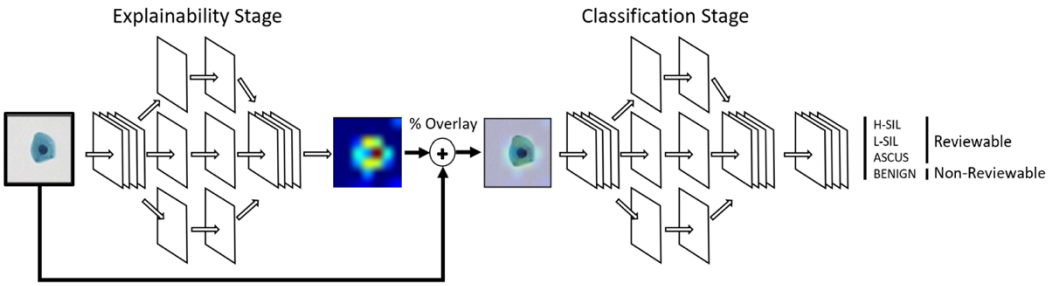


Figura 4.12. Arquitectura propuesta por el modelo compuesto de 2 redes CNN tipo Xception. El modelo de la izquierda permite extraer el mapa de calor que se superpone a la imagen original en función de un porcentaje de ponderación marcado por el parámetro α. Esta imagen fusionada sirve de insumo al modelo de la derecha que clasifica a la célula en una categoría diagnóstica. Elaboración propia.

Con este modelo explicable tratamos de mejorar los resultados de clasificación obtenidos en otras fases del proyecto y analizamos cómo se alteran los resultados con el uso de esta nueva imagen fusionada con el mapa de calor. A través del ajuste de las cifras del parámetro α tratamos de evaluar si hay una mejora de las predicciones de la imagen al enriquecerla con la explicabilidad del modelo. Con el uso de la XIA, esperamos una mejora de los resultados ya que podemos analizar gracias al mapa de calor en qué regiones de la célula se ha fijado el modelo, para así poder estudiar y mejorar de manera más eficiente las predicciones.

4.5.3. BASE DE DATOS EMPLEADA Y SIMPLIFICACIÓN DE LOS RESULTADOS PARA ANALIZAR LA UTILIDAD DEL MODELO

Con el objetivo de mejorar los resultados del modelo, hemos simplificado las 4 categorías diagnósticas del Sistema Bethesda (LSIL, HSIL, ASC-US y benignas) en únicamente en 2 clases, "revisables" y "no revisables". La razón de esta fusión de categorías se debe al desequilibrio entre las clases LSIL y ASC-US, no siendo posible realizar una clasificación en 4 categorías distintas con precisión. Además, existe poca diferenciación entre los mapas de calor de las celdas de LSIL y ASC-US, siendo los mapas de calor en ocasiones superponibles y muy difíciles de diferenciar, incluso sobre la imagen real para patólogos expertos.

De forma práctica, los diagnósticos relevantes de las categorías del Sistema Bethesda, son ASC-US, LSIL y HSIL, ya que conllevan un seguimiento más estrecho de la paciente y; en ocasiones, implica la toma de importantes decisiones terapéuticas como la toma de biopsia cervical para la confirmación histopatológica. Por lo tanto, de forma global, las citologías "revisables" son aquellas con estos diagnósticos relevantes y las "no revisables" pertenecen a aquellas muestras benignas que no requieren un seguimiento especial de la paciente ya que no contienen signos de displasia ni de malignidad. Es así como las células benignas se reclasificarían como "no revisables"; mientras que ASC-US, LSIL y HSIL pasarían a ser "revisables".

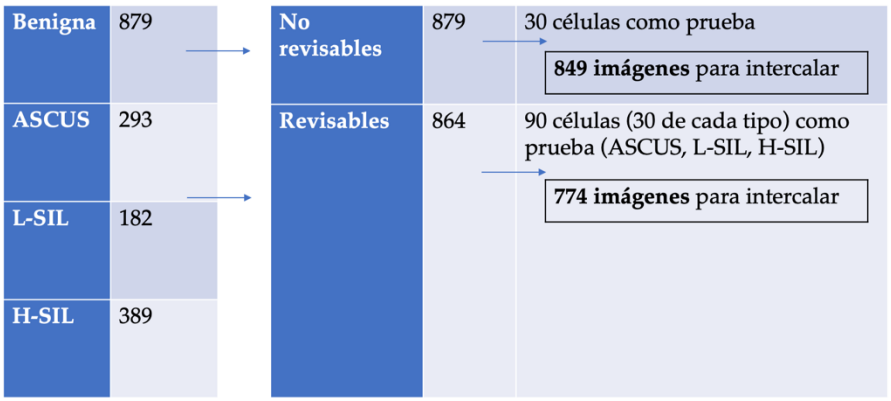
Para llevar a cabo esta fase de estudio hemos partido de nuestras 1405 células reales sobre las que hemos aplicado algunas técnicas de aumento de datos como zoom, rotaciones, desplazamientos, volteos, recortes o distorsiones de la imagen. De esta forma hemos obtenido 1743 imágenes reales de células, no sometidas a ningún procesamiento previo, que fueron clasificadas por un patólogo. Estas células se distribuían en 389 células HSIL, 293 células ASC-US, 182 células LSIL y 879 células benignas. Tras su reagrupación en el sistema binario, nuestra base de datos ha pasado a contener 864 células "revisables" y 879 "no revisables" (**Tabla 4.1**).

Del conjunto total de datos, el 93% (1623 imágenes) se han empleado para intercalar o entrenar al modelo, mientras que el 7% restante (120 imágenes, 30 imágenes de cada subtipo) se ha destinado para probar o testear al modelo. Todas estas imágenes tienen las siguientes funciones:

- Las células para intercalar se emplean para entrenar el modelo, ya que, al ser células previamente clasificadas manualmente, el modelo puede aprender a clasificar.

- Las células de prueba se utilizan para evaluar el modelo y ver si este es capaz de clasificarlas adecuadamente a pesar de no haberse enfrentado nunca a ellas. A través del mapa de calor tratamos de comprender en qué detalles de la célula se ha centrado el modelo para justificar su clasificación.

Tabla 4.1. Base de datos empleada para elaborar el modelo explicable. Elaboración propia.



Por lo tanto, en nuestro modelo se ha usado para prueba un total de 90 imágenes "revisables" y 30 "no revisables"; y para intercalar y entrenar al algoritmo un total de 774 imágenes "revisables" y 849 imágenes "no revisables". Debido al número limitado de imágenes se ha empleado a su vez un generador de datos, que trata de aplicar transformaciones aleatorias de cada imagen; como zoom de hasta el 15%, rotaciones de hasta 10 grados, desplazamientos horizontales o

verticales de hasta el 20%, cortes de hasta el 15% y volteos verticales u horizontales. Al generar estas modificaciones pueden formarse huecos en la imagen que se han solventado mediante el estiramiento del pixel.

Es importante resaltar que estas imágenes de células reales no han sido procesadas previamente al igual que en el resto de fases del estudio. Las imágenes pueden contener áreas plegadas, borrosas y superpuestas, siendo una característica habitual de las células en el diagnóstico por citología, suponiendo un desafío mayor en el diseño y creación de modelos de clasificación de imagen.

Una vez que el modelo ha sido entrenado con estas imágenes no preprocesadas y tras emplear el método Grad-CAM sobre nuestra base de datos, se espera que la red realice un mapa de calor de cada célula, para después emplear nuestra arquitectura híbrida a un problema de clasificación real de células reales fusionadas con los mapas de calor. En la sección de resultados discutiremos qué valor de fusión del parámetro α ofrece mejores resultados, y analizaremos si existe una mejoría en la predicción en el sistema de clasificación bifásica ("revisable" y "no revisables") frente a la clasificación en 4 clases (benigna, ASC-US, LSIL y HSIL).

4.6. EL USO DE LA INTELIGENCIA ARTIFICIAL COMO AYUDA AL DIAGNÓSTICO EN PAÍSES EN VÍAS DE DESARROLLO

El cáncer de cérvix representa el cuarto cáncer más prevalente en mujeres y se trata de la segunda causa de muerte en mujeres en países en vías de desarrollo, únicamente superada por el cáncer de mama. El 85-90% de las muertes por cáncer de cérvix se producen en países en vías de desarrollo con unas tasas de mortalidad 18 veces superiores a las de los países desarrollados. Además, se trata de un cáncer cuyas tasas de supervivencia disminuyen drásticamente si no se diagnostica en las primeras fases de la enfermedad.

El motivo de las cifras tan altas de mortalidad en estos países recae en el difícil acceso al sistema sanitario en aquellos territorios en los que los recursos son limitados. De este modo, las pacientes debutan en estadios muy avanzados de la enfermedad con clínica de dolor o sangrado anómalo en los que el cáncer está avanzado y los tratamientos son agresivos, acarreando una gran morbilidad. En estos entornos de bajos recursos hay escasez de personal sanitario cualificado, incluidos patólogos o citotécnicos capaces de diagnosticar con precisión las citologías cervicovaginales, por lo que es difícil detectar a las pacientes en estadios precoces de la enfermedad. Estas limitaciones se ven agraviadas por barreras geográficas y económicas que limitan el acceso de las pacientes a los servicios sanitarios.

El DL junto a dispositivos de bajo coste se presentan como una solución prometedora para hacer frente a la gran demanda de este servicio diagnóstico de estas pacientes, con un elevado número de citologías y limitada disponibilidad de patólogos cualificados.

4.6.1. PROPUESTA DE UN MODELO DE BAJO COSTE EN PAÍSES EN VÍAS DE DESARROLLO

La principal barrera que encontramos en la implantación de sistemas de IA en Anatomía Patológica es la necesidad de la digitalización de las muestras. Una vez que obtenemos las laminillas citológicas con las células cervicales de la paciente, y tras su tinción con el método de Papanicolau, debemos digitalizarlas para poder aplicarle algoritmos de IA.

El método más comúnmente empleado en nuestro país es el empleo de escáneres de laminillas para la digitalización de muestras. A pesar de que la resolución de las imágenes obtenidas son excelentes, presentan la limitación de su elevado coste; no solo en la compra del equipo, sino en su mantenimiento y la necesidad de disponer de grandes espacios de almacenamiento para todas las imágenes digitales.

En países con recursos limitados, el acceso a estas tecnologías es limitado, por lo que plateamos la utilidad de la aplicación de redes de diagnóstico celular CNN sobre dispositivos de bajo rendimiento, como video-capturadoras o dispositivos de bajo coste como teléfonos móviles.

El inconveniente que plantean estos dispositivos es la mala calidad de la imagen, que puede alterar la capacidad del modelo para clasificar la imagen. De este modo, estudiamos algunas vías de mejora de calidad de la imagen que permitan implementar estas redes CNN como método de clasificación celular sobre imágenes obtenidas de dispositivos de bajo rendimiento.

V – RESULTADOS

V - RESULTADOS

5.1. NUEVA BASE DE DATOS MEDIANTE LA CREACIÓN DE CÉLULAS ARTIFICIALES

Mediante el uso de las redes antagónicas StyleGAN2, hemos generado 2000 células artificiales de cada categoría diagnóstica, aumentando nuestra base de datos de 1405 células reales a un total de 8000 células artificiales (**Figura 5.1** y **Tabla 5.1**). En categorías diagnósticas como benigna en la que disponíamos de una base de datos cuantiosa, hemos cuadriplicado el número de células; mientras que en categorías menos cuantiosas en imágenes como LSIL, hemos multiplicado por 9 dicha cantidad.

Tabla 5.1. Distribución de las imágenes de las células reales obtenidas y las células artificiales generadas, entre las distintas categorías diagnósticas del Sistema Bethesda. Elaboración propia.

	Imágenes reales	Imágenes artificiales
Benignas	450	2000
ASC-US	323	2000
LSIL	213	2000
HSIL	419	2000
Total	1405	8000

En la próxima fase del estudio se testean estas células artificiales para verificar si han sido generadas correctamente, descartando todas aquellas que han sido elaboradas erróneamente por la red. Las células que serán descartadas por su etiquetado erróneo son 1911 células (conservando únicamente las correctas, 6089), tal y como se desarrollará en el siguiente apartado y se expone en la **Figura 5.2** y la **Tabla 5.2.**

5.2. NUESTRO MODELO HÍBRIDO DCGAN EN LA CLASIFICACIÓN CÉLULAS ESCAMOSAS CERVICALES

Mediante el modelo híbrido propuesto DCGAN se han puesto a prueba las células artificiales generadas por las redes antagónicas a través de la predicción automática por la red, comprobando si el algoritmo DL las etiqueta en la categoría para la que han sido creadas, verificando así que se han generado correctamente, siendo superponibles a la calidad de las células reales (**Figura 5.1**).

En la **Tabla 5.2** hemos elaborado una matriz de confusión con los resultados de la clasificación por la red, enfrentando las categorías iniciales de las células artificiales para las que fueron creadas (2000 para cada categoría, dispuestas en las filas de la matriz de confusión) con la clasificación automática del modelo híbrido (detalladas en las columnas), para así comparar el grado de acierto en el etiquetado celular de cada categoría diagnóstica y ver la autenticidad de las células artificiales generadas. Analizando la matriz de confusión podemos analizar; por ejemplo, del total de 2000 células artificiales HSIL que se han creado, cuantas células la red etiqueta como HSIL (1752), cuántas como benignas (19), cuántos ASC-US (127) o cuántos LSIL (102).

Tabla 5.2. Matriz de confusión en la que se comparan los resultados reales generados por un patólogo experto con los obtenidos de forma automática por el modelo DCGAN. Las filas determinan los diagnósticos para los que la célula artificial fue generada (2000 por cada categoría diagnóstica) y en las columnas la etiqueta generada por el modelo. Elaboración propia.

	PREDICCIÓN AUTOMÁTICA				
Célula artificial	*Benigna*	*HSIL*	*ASC-US*	*LSIL*	***Total***
Benigna	1773	143	78	6	2000
H-SIL	19	1752	127	102	2000
ASC-US	45	102	1426	427	2000
L-SIL	104	173	585	1138	2000

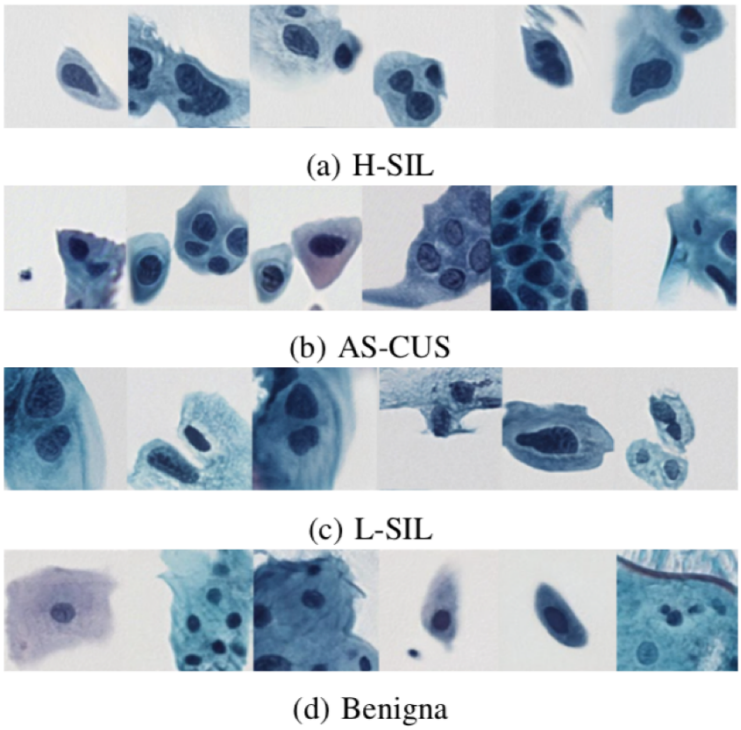


Figura 5.1. Imágenes generadas por la GAN para cada categoría diagnóstica: (a) HSIL, (b) ASC-US, (c) LSIL y (d) benigna. Elaboración propia.

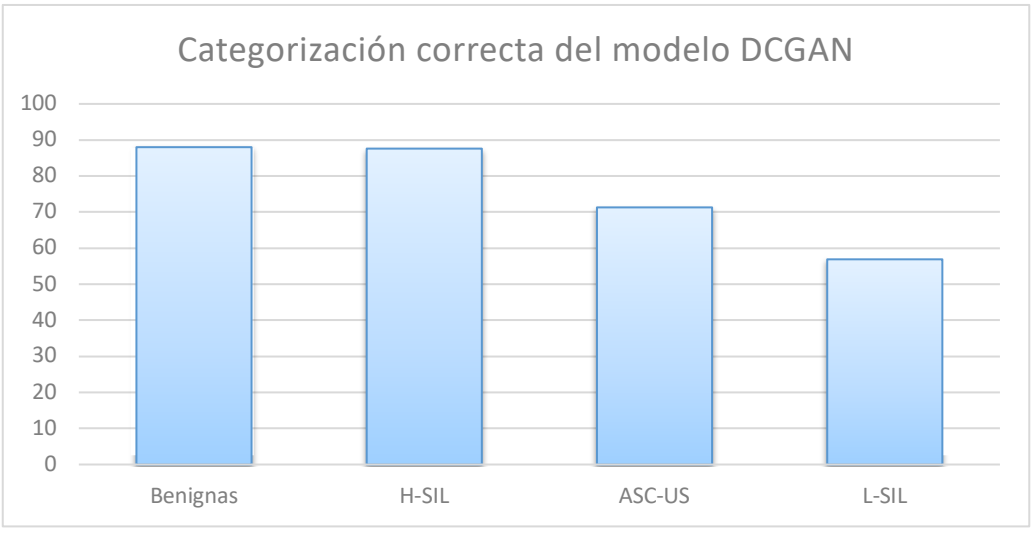


Figura 5.2. Porcentaje de acierto en la elaboración y etiquetado de células artificiales del modelo DCGAN de cada categoría diagnóstica. Elaboración propia.

La **Figura 5.2** permite estudiar de manera visual la distribución entre las células correctamente etiquetadas de cada categoría; es decir, la sensibilidad del modelo para etiquetar la célula en la categoría para la que fue creada dando un resultado verdadero positivo. En la tarea de clasificación de las células benignas el modelo lo hizo correctamente con 1773 células del total de 2000 (que representa el 88,65% del total de células generadas), clasificándose erróneamente 143 como HSIL, 78 como ASC-US y únicamente 6 como LSIL. Respecto a las células HSIL, el modelo DCGAN creó y clasificó correctamente el 87,6% de las 2000 células artificiales de su categoría (1752), en la categoría ASC-US acertó con 71,3% (1426) y en LSIL con 56,9% (1138) (**Tabla 5.2 y 5.3**).

Se realiza un estudio de las métricas por clase analizando los datos obtenidos en la matriz de confusión de la **Tabla 5.2** para analizar la fiabilidad del modelo, estudiando: los verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN) y los verdaderos negativos (VN), que se exponen en la **Tabla 5.3.** Las categorías benigna y HSIL presentan una tasa de VP similar (1773 y 1752 células, respectivamente) y de FN bastante próxima (con 227 y 248 células, respectivamente). En la tabla destacan en la columna de FP la categoría ASC-US con 790 casos y en la columna FN LSIL con 862.

Tabla 5.3. Tasa de verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN) y falsos negativos (FN) obtenidos por el modelo de clasificación de células artificiales. Elaboración propia.

Clase	*VP*	*FP*	*VN*	*FN*
Benigna	1773	168	5832	227
HSIL	1752	418	5582	248
ASC-US	1426	790	5210	574
LSIL	1138	535	5465	862

Del mismo modo, en la **Tabla 5.4** se detallan la sensibilidad (S), especificidad (E), precisión (P), valor predictivo negativo (VPN), exactitud (ACC del inglés, *accurancy*), valor F1, tasa de falsos positivos (TFP) y tasa de falsos negativos (TFN); que se definen a continuación:

- Sensibilidad o tasa de verdaderos positivos: mide el grado en que el modelo identifica correctamente los elementos de una clase. Se calcula con la fórmula:
  Sensibilidad = VP / (VP + FN).

- Especificidad determina la capacidad del modelo en detectar los casos negativos reales detectados correctamente por la red. Se calcula de la siguiente forma:
  Especificidad = VN / (VN + FP).

- Precisión: es la proporción del número de elementos que realmente pertenecen a una clase, entre el número total de elementos de esa clase, incluidas las predicciones del modelo. La fórmula para calcular la precisión es:
  Precisión = VP / (VP + FP).

- Valor predictivo negativo: trata la proporción de predicciones negativas que son correctas; es decir, que realmente son negativas. Se calcula:
  VN / (VN + FN)

- Exactitud: es la proporción del número total de predicciones correctas respecto al número total de predicciones. Se calcula utilizando la siguiente fórmula: Exactitud = (VP+VN) / (VP + VN + FP + FN).

- Valor F1: combina precisión y sensibilidad para obtener un rendimiento global. Métrica que indica el rendimiento del modelo. Su fórmula es:
  F1 = 2 x (precisión x sensibilidad) / (precisión + sensibilidad).

- Tasa de falsos positivos: trata la proporción de casos negativos que han sido clasificados erróneamente como positivos. Se calcula de la siguiente forma:
  FP / (FP + VN).

- Tasa de falsos negativos: se refiere a la proporción de casos positivos reales que se han clasificado erróneamente como negativos. Su fórmula es:
  FN / (FN + VP).

Tabla 5.4. Análisis de métricas de sensibilidad (S), especificidad (E), precisión (P), valor predictivo positivo (VPN), exactitud (del inglés accurancy, ACC), valor F1 (F1), tasa de falsos positivos (TFP) y tasa de falsos negativos (TFN). Elaboración propia.

Clase	*S (%)*	*E (%)*	*P (%)*	*VPN (%)*	*ACC (%)*	*F1 (%)*	*TFP (%)*	*TFN (%)*
Benigna	88.65	97.20	91.34	96.25	95.06	89.98	2.80	11.35
HSIL	87.60	93.03	80.74	95.75	91.68	84.03	6.97	12.40
ASC-US	71.30	86.83	64.35	90.08	82.95	67.65	13.17	28.70
LSIL	56.90	91.08	68.02	86.38	82.54	61.97	8.92	43.10

En la **Tabla 5.4** podemos ver la comparativa de cada métrica por categoría diagnóstica. Los porcentajes más elevados corresponden a la clasificación benigna, con una S de 88,65%; E de 97,20%; P de 31,34%; VPN del 96,25%, exactitud (en la tabla ACC) de 95,06% y valor F1 del 89,98%. Estos resultados los sigue de cerca con unas cifras muy cercanas, la categoría HSIL. La categoría ASC-US destaca por su alta TFP (13,17%) y LSIL por su elevada TFN (43,10%).

El modelo presenta distintas cifras de sensibilidad dependiendo de la categoría analizada, tal como queda reflejado en la **Figura 5.2, Tabla 5.4** y la **Tabla 5.5.** Se detalla; del total de las 2000 células artificiales generadas para una categoría, cuantas fueron etiquetadas correctamente por la red. En la **Tabla 5.5** se marca en verde la casilla que refleja las predicciones correctas y representa el porcentaje de células que la red ha elaborado y etiquetado correctamente, que se conservan y se reservan para estudio.

Tabla 5.5. Tasa de predicción del modelo DCGAN de las 2000 células artificiales generadas para cada categoría diagnóstica. Se detalla el porcentaje de células artificiales clasificada en cada categoría diagnóstica por el modelo. Elaboración propia.

	PREDICCIÓN DEL MODELO DCGAN			
Célula artificial	*Benigna*	HSIL	ASC-US	LSIL
Benigna	88,65%	7,15%	3,9%	0,30%
HSIL	0,95%	87,6%	6,35%	5,10%
ASC-US	2,25%	5,1%	71,3%	21,35%
LSIL	5,2%	8,65%	29,25%	56,9%

El resto de células sobrantes se descarta por haber sido elaboradas o etiquetadas erróneamente, que alcanza el 11,35% de las células artificiales benignas, el 12,4% de las HSIL, el 28,7% de los ASC-US y el 43,1% de los LSIL. Del total de 8000 células artificiales (2000 por cada categoría), se descartan las 1911 células mal clasificadas y se conservan 6089 células clasificadas correctamente. Estas cifras se obtienen tras sumar, respectivamente, las células mal y bien etiquetadas de todas las categorías diagnósticas, tal y como se detalla en la matriz de confusión de la **Tabla 5.2.**

Tras la reagrupación de los resultados del modelo en una clasificación binaria, hemos elaborado las **Tablas 5.6 y 5.7.** La clasificación binaria se distribuye en dos categorías en lugar de 4; tras reagrupar las células en "no revisables" (que incluyen las células benignas) y "revisables" (que engloban las ASC-US, LSIL y HSIL). La **Tabla 5.6** se trata de una matriz de confusión que nos permite ver en las filas las predicciones reales de las células frente a las columnas que representan las predicciones del algoritmo. La red ha etiquetado correctamente 1773 células como "no revisables" (del total de 2000) y 5832 células como "revisables" (del total de 6000 de este grupo, correspondientes con el sumatorio de las 2000 células de las categorías ASC-US, LSIL y HSIL). Estos datos se corresponden con una sensibilidad del 89% para las células "no revisables" frente al 97% de las "revisables" (**Tabla 5.7**).

Tabla 5.6. Matriz de confusión tras la reclasificación de los resultados en "revisables" (ASC-US, LSIL y HSIL) y "no revisables"(benignas) comparando las predicciones de la red frente a la clasificación real. En las filas se detalla la clasificación real y en las columnas la predicción automática del modelo. Elaboración propia.

	PREDICCIÓN AUTOMÁTICA		
REAL	*No revisable*	*Revisable*	*Total*
No revisable	1773	227	2000
Revisable	168	5832	6000

Tabla 5.7. Porcentaje de células etiquetadas como "revisables" y "no revisables" tras su reclasificación binaria del total de 2000 células "no revisables" y 6000 "revisables". Elaboración propia.

	PREDICCIÓN AUTOMÁTICA	
REAL	*No revisable*	*Revisable*
No revisable	89%	11%
Revisable	3%	97%

5.3. LA SEGMENTACIÓN COMO MEJORA DEL MODELO DE DIAGNÓSTICO CELULAR

5.3.1. ELABORACIÓN AUTOMÁTICA DE MÁSCARAS DE SEGMENTACIÓN

Para entrenar a la red hemos segmentado manualmente las células, delimitado de forma manual el núcleo y citoplasma utilizando distintos colores y despreciando el resto de elementos de fondo de la imagen, para así mostrar con mayor evidencia las irregularidades de la forma y contorno del núcleo y las alteraciones en la relación núcleo-citoplasma; hallazgos clave en el diagnóstico citológico. Hemos marcado en la segmentación los núcleos de rojo y los citoplasmas de azul **(Figura 5.3).**

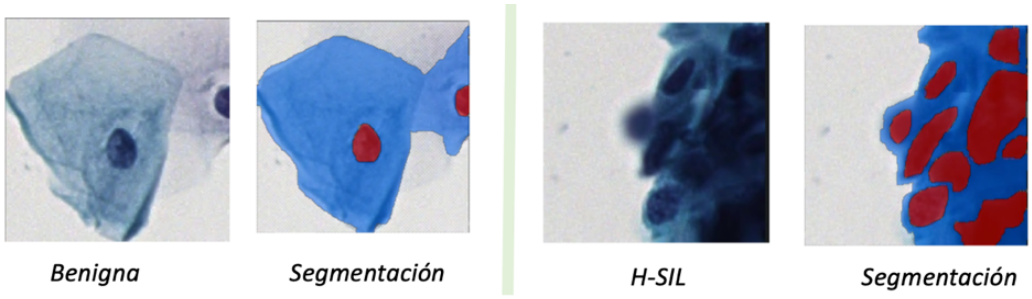


Figura 5.3. Imágenes que comparan la segmentación manual con la imagen real. A la izquierda una célula benigna y a la derecha una célula H-SIL con sus respectivas segmentaciones, marcando en rojo los núcleos y en azul los citoplasmas. Elaboración propia.

Una vez hemos obtenido nuestras células segmentadas, hemos entrenado al algoritmo DL con el 5% de nuestra muestra de células segmentadas para testear si es capaz de segmentar de manera automática nuevas imágenes a las que hemos enfrentado la red. En la **Figura 5.4** se muestran los resultados de la segmentación automática obtenidos por la U-Net, en la que se observa que la red es capaz de discernir claramente entre el núcleo, citoplasma y fondo (que marca en colores rojo, verde y azul, respectivamente).

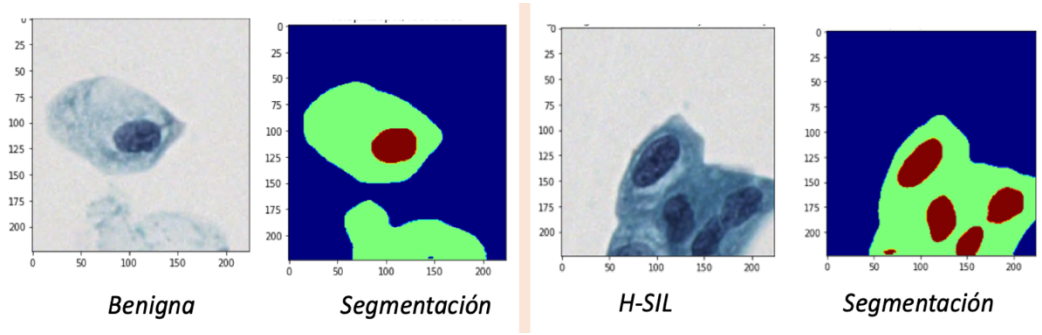


Figura 5.4. Resultado de la segmentación automática mediante algoritmos de deep learning. A la izquierda una célula benigna y a la derecha HSIL, diferenciando el núcleo en rojo, el citoplasma en verde y el fondo el azul. Elaboración propia.

5.3.2. CLASIFICACIÓN CELULAR TRAS LA SEGMENTACIÓN

A continuación, falta comparar cómo se comporta la red ante una nueva fase de clasificación una vez ha aprendido a segmentar automáticamente las imágenes; de modo que mediante el uso de una imagen segmentada en la entrada a la red, el algoritmo sabrá en qué zonas de la célula ha de fijarse para emitir su diagnóstico. Para ello y tal como se expuso en el apartado de material y métodos, se han evaluado 3 modelos distintos de los cuales uno no realiza una segmentación preliminar y los otros dos sí, diferenciándose entre sí únicamente por el método de fusión de redes o clasificador empleado. Estos modelos son los siguientes (**Figura 5.5)**:

- Un *modelo clásico* Xception que no emplea métodos de segmentación previos a la clasificación.

- *Modelo 1* que sí realiza una segmentación preliminar celular mediante el empleo de una red U-Net para su posterior clasificación por una red Xception que emite un diagnóstico final. La fusión entre ambas redes en este caso se realiza mediante un clasificador que emplea una capa densa y una capa aplanada.

- *Modelo 2* que también realiza una segmentación empleando la misma arquitectura que el *modelo 1* pero que emplea como clasificador una capa GAP para la fusión de las redes.

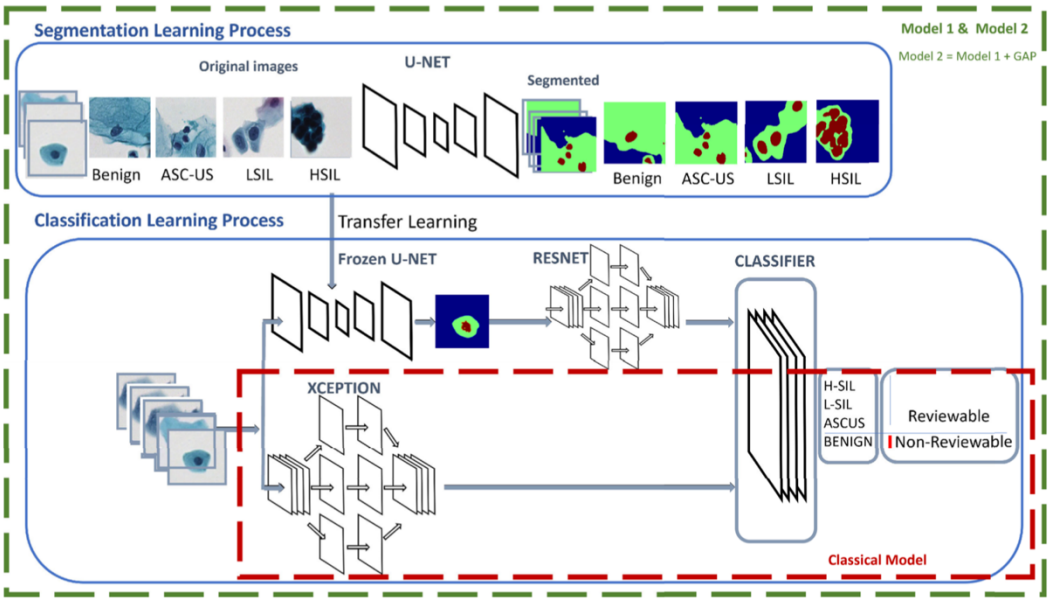


Figura 5.5. Modelos propuestos, donde la U-Net aprende a segmentar y se fusiona con una red Xception mediante un clasificador. Marcado en verde el contorno de la arquitectura de los modelos 1 y 2 y en rojo del modelo clásico. Elaboración propia (Ortiz-González et al., 2025).

Con estos modelos haremos un análisis de resultados mediante el estudio de distintas métricas; tanto para la clasificación multiclase en 4 categorías diagnósticas (benigna, ASC-US, LSIL y HSIL), como para la clasificación binaria en 2 clases ("revisable" y "no revisable"). A priori, la clasificación en 2 clases es más práctica ya que sirve como filtro para que el patólogo decida si debe revisar o no la citología, de manera que solo es necesario que revise aquellas citologías con riesgo de tener displasia clasificadas como "revisables", marcando desde el diagnóstico un seguimiento clínico más estrecho para estas pacientes.

Para evaluar el rendimiento de los modelos se han seleccionado métricas que ofrecen información sobre las fortalezas y limitaciones de los mismos. Estas métricas son: *exactitud, precisión, sensibilidad, valor F1 y área bajo la curva*. La mayoría

de estos parámetros se han estudiado anteriormente en otras fases del estudio. En este caso añadimos el estudio del área bajo la curva ROC (del inglés, *AUC*) que representa la probabilidad de que el modelo, dando un ejemplo positivo y uno negativo elegidos aleatoriamente, clasifique el positivo por encima del negativo.

5.3.2.1. ***LA CLASIFICACIÓN MULTICLASE PARA LOS MODELOS PROPUESTOS***

La **Figura 5.6** se trata de un gráfico de barras que presenta la comparación de los resultados de las métricas de la clasificación multiclase. Se representa cada modelo con una barra de un color diferenciado y se evalúan distintas métricas: el área bajo la curva (en la figura, AUC), valor F1 (F1 score), sensibilidad (en la figura se muestra como *recall*, del inglés), precisión (precision), y exactitud (en la figura *accurancy*, del inglés). El *modelo 2* presenta unas cifras para *AUC* de 0,919*; F1score* 0,727; *recall* 0,725*; precisión* 0,747 *y accurancy* 0,725. El *modelo 1* presenta unos resultados similares con unas cifras para *AUC* de 0,909*; F1score* 0,702; *recall* 0,708*; precisión* 0,712 *y accurancy* 0,708. Dichas métricas se pueden comparar con el modelo clásico con unas cifras para *AUC* de 0,866*; F1score* 0,666; *recall* 0,675*; precisión* 0,682 *y accurancy* 0,675.

Los *modelos 1 y 2* han presentado a priori unas métricas muy parecidas. Para tratar de completar el estudio, comparamos los resultados de la clasificación multiclase entre modelos con el *test de Wilcoxon*, con el cual observamos que:

- Al comparar el modelo clásico con el modelo 1, se obtiene una $p = 0.1$, diferencia estadísticamente no significativa.

- Al comparar el *modelo clásico* con el *modelo 2*, se obtiene una $p=0,05$, diferencia estadísticamente significativa.

- Al comparar el *modelo 1 y 2* no existen diferencias estadísticamente significativas.

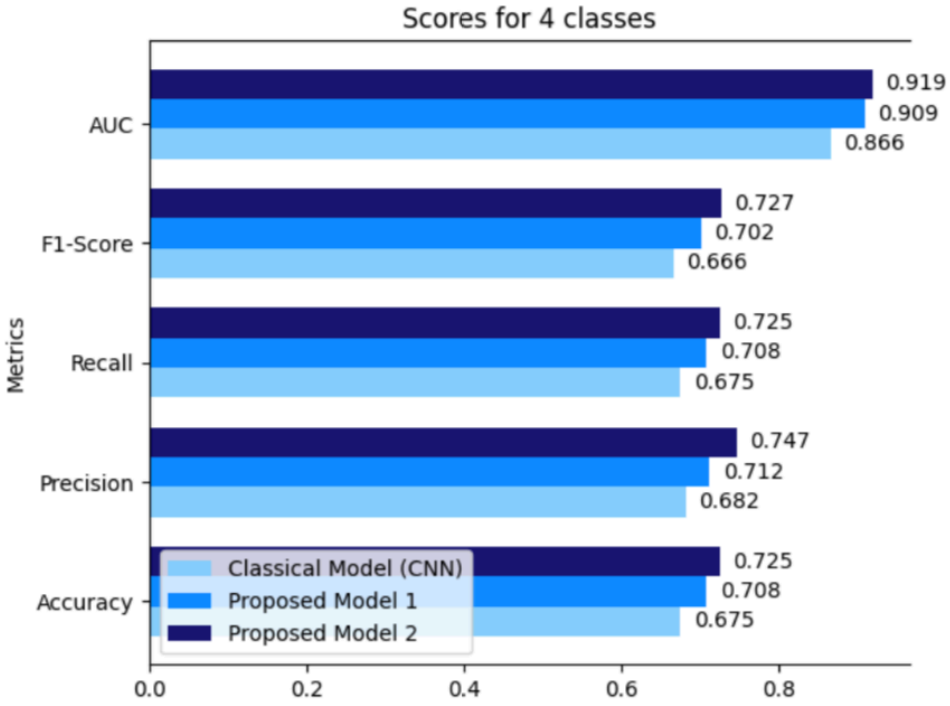


Figura 5.6. Resultados de la clasificación del modelo en 4 clases, comparando el modelo propuesto 1, 2 y el modelo clásico de referencia. Del inglés, "AUC" hace referencia al área bajo la curva, "F1-score" al valor F1, "recall" a la sensibilidad o tasa de verdaderos positivos, "precision" a precisión y "accuracy" a exactitud. Elaboración propia (Ortiz-González et al., 2025).

5.3.2.2. LA CLASIFICACIÓN BINARIA PARA LOS MODELOS PROPUESTOS

Se presenta en la **Figura 5.7** un gráfico de barras que muestra la comparación de los resultados de las métricas de la clasificación binaria. Del mismo modo que en el caso anterior, los distintos modelos se presentan con una barra de color diferenciado. El *modelo 2* presenta unas cifras para *AUC* de 0,939*; F1score* 0,943; *recall* 0,942*; precisión* 0,945 *y accurancy* 0,942. El *modelo 1* presenta unas cifras para *AUC* de 0,933*; F1score* 0,935; *recall* 0933*; precisión* 0,938 *y accurancy* 0,933. También se presentan las métricas del modelo clásico con unas cifras para *AUC* de 0,878*; F1score* 0,858; *recall* 0,85*; precisión* 0,889 *y accurancy* 0,85.

Cuando comparamos estos resultados de la clasificación binaria de los modelos entre sí a través de la prueba de Wilcoxon, encontramos que:

- Al comparar el *modelo clásico* con el *modelo 1* obtenemos una p=0,03, diferencia estadísticamente significativa.

- Al comparar el *modelo clásico* con el *modelo 2* obtenemos una p=0,01, diferencia estadísticamente significativa.

- De nuevo, al comparar el *modelo 1 y 2* en la clasificación binaria no existen diferencias estadísticamente significativas.

La **Tabla 5.8** es una tabla comparativa de los resultados obtenidos con cada modelo en la clasificación multiclase y en la clasificación binaria; distribuidos en las columnas verde y azul, respectivamente, distribuyéndose en las filas cada uno de los parámetros estadísticos estudiados anteriormente. Nos permite ver que las cifras más altas se obtienen en la clasificación binaria, concretamente en el *modelo 2* con AUC de 0.939, valor F1 de 0.943, sensibilidad (recall) de 0.942, precisión de 0.945 y exactitud de 0.942; frente a las métricas del modelo clásico con 0.878 AUC, 0.858 valor F1, 0.850 sensibilidad, precisión 0.889 y exactitud 0.850.

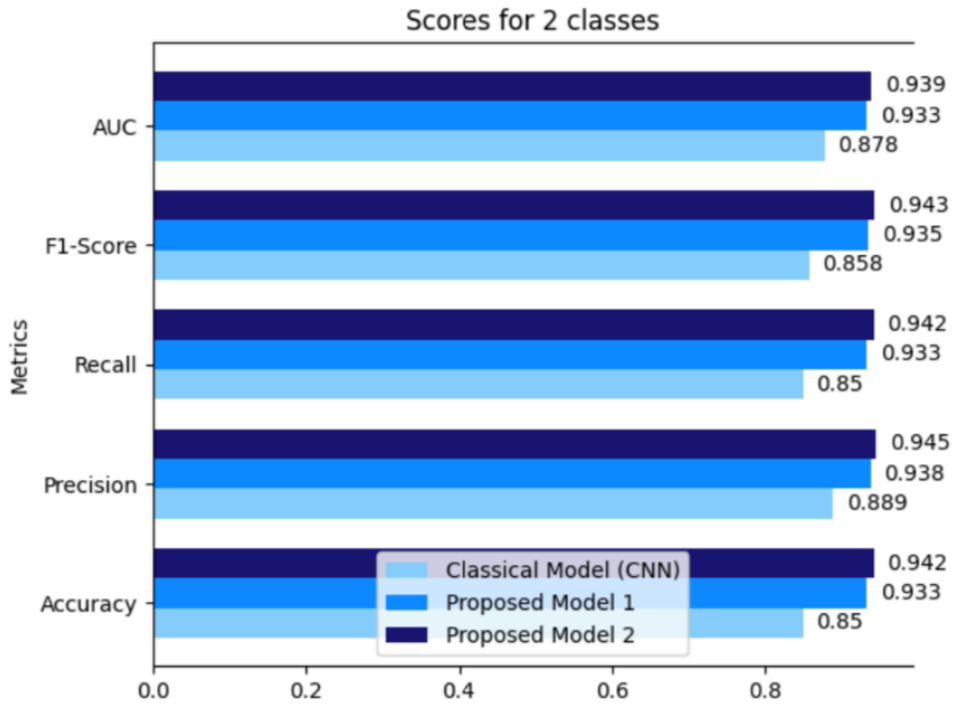


Figura 5.7. Resultados de la clasificación del modelo en 2 clases en el que se comparan las métricas para los 3 modelos a estudio. Del inglés, "AUC" hace referencia al área bajo la curva, "F1-score" al valor F1, "recall" a la sensibilidad o tasa de verdaderos positivos, "precision" a precisión y "accuracy" a exactitud. Elaboración propia (Ortiz-González et al., 2025).

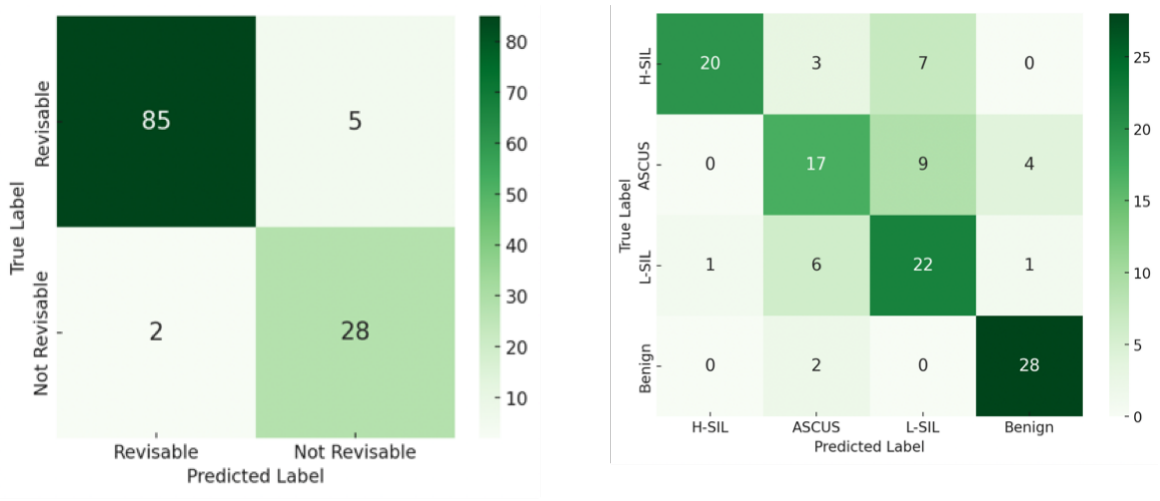
Tabla 5.8. Tabla comparativa de las métricas de los 3 modelos estudiados, enfrentando la clasificación multiclase con la binaria. Elaboración propia.

Métrica	CLASIFICACIÓN MULTICLASE			CLASIFICACIÓN BINARIA		
	Modelo Clásico	*Modelo 1*	*Modelo 2*	*Modelo Clásico*	*Modelo 1*	*Modelo 2*
AUC	0.866	0.909	0.919	0.878	0.933	**0.939**
F1-score	0.666	0.702	0.727	0.858	0.935	**0.943**
Recall	0.675	0.708	0.725	0.850	0.933	**0.942**
Precisión	0.682	0.712	0.747	0.889	0.938	**0.945**
Accuracy	0.675	0.708	0.725	0.850	0.933	**0.942**

5.3.2.3. ***COMPARATIVA DE LA CLASIFICACIÓN MULTICLASE Y BINARIA***

Una vez hemos comparado todos los modelos, hemos seleccionado el modelo con el que hemos obtenido mejores métricas; el *modelo 2* (con su capa GAP como clasificador) para someterlo a una nueva fase de clasificación con 30 células de cada categoría diagnóstica. A continuación, hemos elaborado una matriz de confusión; tanto multiclase como binaria, para ver si se producen modificaciones importantes en los resultados obtenidos por la red del *modelo 2*. La matriz se refleja en la **Figura 5.8**, donde las filas representan las etiquetas verdaderas y las columnas las etiquetas predichas por el modelo (Ortiz-González et al., 2025). Cada celda muestra el número de instancias para cada combinación de etiquetas verdaderas y predichas, lo que permite el cálculo de la precisión del modelo. La diagonal muestra el número de instancias clasificadas correctamente, mientras que los datos de fuera de la diagonal presentan errores de clasificación.

De este modo podemos ver en la clasificación multiclase (**Figura 5.8-b**) de las 30 células testeadas de cada categoría diagnóstica (con una muestra total de 120 células) cuántas fueron clasificadas correctamente. Obtenemos una clasificación correcta en 20 casos de HSIL, 17 de ASC-US, 22 de LSIL y 28 de benignas. En el resto de celdillas podemos ver como se distribuyen los diagnósticos erróneos; por ejemplo, en HSIL la mayoría de errores diagnósticos correspondieron a casos etiquetados como LSIL (7 casos), en ASC-US a casos etiquetados como LSIL (9 casos), en LSIL a casos etiquetados como ASC-US (6 casos) y en benignas a casos etiquetados como ASC-US (únicamente en dos casos).



a. Clasificación en 2 clases

b. Clasificación en 4 clases

Figura 5.8. Matriz de confusión del mejor modelo propuesto (modelo 2) para analizar el rendimiento del modelo en una clasificación binaria (a) y multiclase en 4 categorías diferentes (b). Elaboración propia (Ortiz-González et al., 2025).

En la **Tabla 5.9** se detallan las métricas obtenidas tras el estudio de la matriz de confusión de la clasificación multiclase, reflejadas en la **Figura 5.8-b**. Los casos verdaderos positivos (VP) hacen referencia a los casos positivos detectados por la red; es decir, a la sensibilidad del estudio sobre cada categoría de una muestra total de 120 células (30 de cada categoría diagnóstica). En la fila de la sensibilidad se puede apreciar que el valor mayor lo presentan las células benignas con un 93,33%; en la especificidad las HSIL con un 98,89%, en la precisión las HSIL con un 95,24%, en el valor predictivo negativo en las células benignas con un 97,70%, en la exactitud las células benignas con un 94,17% y en el valor F1 las células benignas con un 88,89%.

Tabla 5.9. Tabla de las métricas obtenidas tras el análisis de los resultados de la matriz de confusión de la clasificación multiclase del modelo 2. La muestra consta de 30 células en cada categoría, con un total de 120. Se estudian: verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN), verdaderos negativos (VN), sensibilidad, especificidad, precisión , valor predictivo negativo, exactitud, valor F1, tasa de falsos positivos y tasa de falsos negativos. Elaboración propia.

Métricas Clasificación multiclase	**HSIL**	**ASC-US**	**LSIL**	**Benigna**
VP	20	17	22	28
FP	1	11	16	5
FN	10	13	8	2
VN	89	79	74	85
Sensibilidad	66,67 %	56,67 %	73,33 %	93,33 %
Especificidad	98,89 %	87,78 %	82,22 %	94,44 %
Precisión	95,24 %	60,71 %	57,89 %	84,85 %
Valor predictivo negativo	89,90 %	85,87 %	90,24 %	97,70 %
Exactitud	90,83 %	80 %	80 %	94,17 %
F1	78,43 %	58,62 %	64,71 %	88,89 %
Tasa de falsos positivos	1,11 %	12,22 %	17,78 %	5,56 %
Tasa de falsos negativos	33,33 %	43,33 %	26,67 %	6,67 %

Con la matriz de confusión de la clasificación binaria de la Figura **5.8-a**, podemos obtener las métricas estadísticas de la clasificación binaria, que se reflejan en la **Tabla 5.10.** Aquí hay que destacar que la muestra se distribuye de manera distinta a la clasificación multiclase. Nuestra muestra de 120 células repartidas en 30 células de cada categoría diagnóstica multiclase, en la clasificación binaria se agrupa únicamente en 2 categorías: "revisables" que incluyen la suma de ASC-US, LSIL y HSIL; y la categoría "no revisables" que engloba únicamente las células

benignas. Por ello, la categoría "revisables" contiene 90 células y la "no revisable" 30 células. En la sensibilidad destaca la categoría "revisable" y en la especificidad la "no revisable", ambas con un 94,44%; en la precisión destacan las "revisables" y en el valor predictivo negativo las "no revisables" ambas con un 97,7%.

Tabla 5.10. Tabla con las métricas obtenidas tras el análisis de la matriz de confusión de la clasificación celular binaria del modelo 2. La muestra consta de 90 células revisables y 30 no revisables; alcanzando una muestra total de 120 células. Se estudian: verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN), verdaderos negativos (VN), sensibilidad, especificidad, precisión , valor predictivo negativo, exactitud, valor F1, tasa de falsos positivos y tasa de falsos negativos. Elaboración propia.

Métricas Clasificación binaria	**Revisable**	**No revisables**
VP	85	28
FP	2	5
FN	5	2
VN	28	85
Sensibilidad	94,44 %	93,33 %
Especificidad	93,33 %	94,44 %
Precisión	97,70 %	84,85 %
Valor predictivo negativo	84,85 %	97,70 %
Exactitud	94,17 %	94,17 %
F1	96,05 %	88,89 %
Tasa de falsos positivos	6,67 %	5,56 %
Tasa de falsos negativos	5,56 %	6,67 %

Tabla 5.11. Distribución de la clasificación celular comparando el sistema binario y el multiclase del modelo 2. Elaboración propia.

	PREDICCIÓN BINARIA	
REAL	*No revisable*	*Revisable*
No revisable	93%	6%
Revisable	7%	94%

	PREDICCIÓN MULTICLASE	
REAL	*Benignas*	*ASC-US + LSIL + HSIL*
No revisable	93%	35%
Revisable	7%	65%

La **Tabla 5.11** nos muestra una comparativa del porcentaje de aciertos y errores de cada tipo de clasificación, que nos permite compararlas entre sí. La **Tabla 5.12** nos permite ver la distribución de los diagnósticos fallidos en la clasificación multiclase, destacando que hasta el 30% de los ASC-US se diagnosticaron como LSIL (9 casos), el 20% de los LSIL como ASC-US (6 casos) y 23,33% de las HSIL fueron diagnosticadas erróneamente como LSIL (7 casos) (**Tabla 5.12** y **Figura 5.8-b**).

Tabla 5.12. Porcentaje de células clasificadas correctamente en cada categoría de la predicción multiclase del modelo 2. Nótese en las celdillas marcadas en verde las células correctamente clasificadas, que representan la sensibilidad de la red para cada categoría. Elaboración propia.

	PREDICCIÓN MULTICLASE			
REAL	*HSIL*	ASC-US	LSIL	Benigna
HSIL	67%	10%	23%	0%
ASC-US	0%	56%	30%	13,33%
LSIL	3,33%	20%	73,3%	3,33%
Benigna	0%	6,66%	0%	93,33%

5.4. EL IMPACTO DE LA EXPLICABILIDAD EN LA CLASIFICACIÓN CELULAR

La arquitectura híbrida del modelo muestra una etapa de explicabilidad seguida de una etapa de clasificación celular. En la etapa de explicabilidad la red aprende a generar mapas de calor con los que elaboran imágenes de células enriquecidas con mapas de calor, para pasar a una segunda etapa de clasificación con la que se espera una mejora de los resultados de clasificación con estas células enriquecidas como imagen de entrada.

5.4.1. Primera etapa de explicabilidad

En esta etapa de explicabilidad se emplea una red CNN tipo Xception preentrenada que se combina con una Grad-CAM como método de explicabilidad para obtener un mapa de calor completo de la imagen celular.

A través de pequeñas modificaciones de la imagen el algoritmo es capaz de distinguir que zonas de la imagen son más importantes en la clasificación celular, ya que al realizar pequeñas modificaciones en estas regiones relevantes se producen grandes cambios en la información de salida de la clasificación celular.

De este modo, la red genera un mapa de calor en el que marca en rojo las zonas más importantes de la imagen y en tonos azules las menos relevantes. En los mapas de calor generados por el algoritmo **(Figura 5.9)** vemos cómo marca como zona relevante el núcleo de la célula dándole el mayor peso en la decisión del diagnóstico, de manera superponible a lo que haría un patólogo experto.

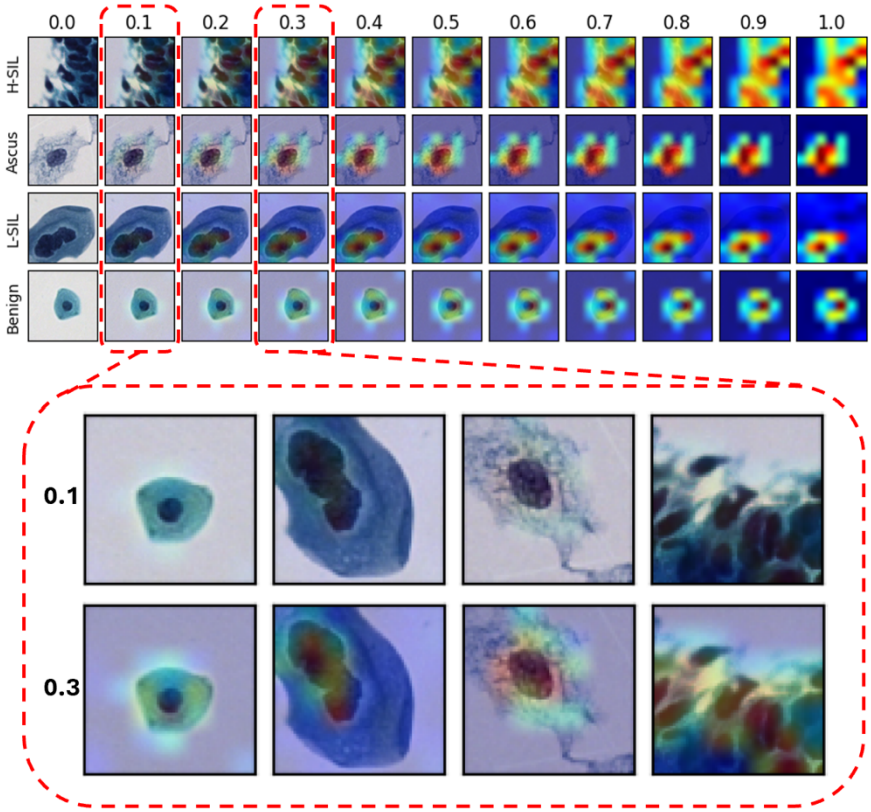


Figura 5.9. Resultados tras la fusión de las imágenes reales con los mapas de calor para distintos valores del parámetro α. Se resaltan los valores de α = 0.1 y 0.3 para ver con más detalle la influencia del mapa de calor. Elaboración propia.

Con todos los mapas de calor, se obtiene una nueva base de datos de imágenes de células fusionadas con el mapa de calor tratando de conservar el equilibrio entre la información que proporciona el mapa de calor señalando las partes más relevantes de la célula, con la información de la célula original. Este grado de fusión de imágenes lo determina el parámetro α y servirá como imagen de entrada a la segunda etapa de clasificación del modelo.

Buscamos cual es el mejor ajuste del parámetro α para el funcionamiento del modelo. Para ello, se realizan varios experimentos probando distintos grados de transparencia del parámetro α de los mapas de calor, que van de 0 a 1, donde 0.0 representa la imagen real y 1.0 únicamente el mapa de calor, abarcando todas las posibilidades intermedias. En la **Figura 5.9** se muestran las imágenes fusionadas

con distintos valores del parámetro α para las categorías benigna, LSIL, ASC-US y HSIL. También se muestra unas celdas expandidas para α =0.1 y α=0.3; donde se pone en evidencia que con el valor 0,3 el mapa de calor diluye partes relevantes de la célula original.

En la **Figura 5.9** podemos ver cómo para células benignas el mapa de calor fusionado con un valor de $\alpha = 0.1$, se centra principalmente en el citoplasma celular y no tanto en el núcleo debido a que marca principalmente el citoplasma celular. En este tipo de células la relación núcleo-citoplasma es baja resultando en núcleos pequeños que el algoritmo apenas marca en rojo, siendo esta una de las características típicas de este tipo de células.

En el caso de las células LSIL, estas tienen unas características celulares propias secundarias a la infección por HPV, como son el aumento de la relación núcleo-citoplasma y la presencia de halo perinuclear. Al emplear un valor de α = 0.1 estas características se ven reflejadas. Sin embargo, cuando empleamos un grado α superior, como $\alpha = 0.3$, estas características nucleares de la imagen original quedan diluidas por el mapa de calor y la red toma como núcleo tanto el propio núcleo como el halo perinuclear, falseando el tamaño verdadero del núcleo. Por otro lado, si empleamos un valor de α más bajo como α = 0.1 la red distingue perfectamente el halo perinuclear del núcleo sin verse diluido por el mapa de calor.

5.4.2. Segunda etapa de clasificación

Una vez la red aprende a generar mapas de calor, pasamos a la etapa de clasificación. Con nuestra imagen enriquecida resultado de fusionar la célula original con su mapa de calor, pasamos a someter a la red a una fase de clasificación. Estudiando los resultados de la clasificación obtenidos, podremos ajustar el mejor valor del parámetro α; es decir, el grado de fusión óptimo, buscando un equilibro entre la información que aporta el mapa de calor y la no pérdida de la información que da la imagen original.

Una vez obtenidas todas las combinaciones de fusión entre mapa de calor y célula original a través de distintos valores de α, se ha ejecutado la CNN Xception 10 veces de manera independiente para cada valor α; desde α=0 hasta α=1.0, para tratar de identificar el mejor valor de α que se ajusta a un mejor etiquetado celular en la salida de la red. Como imagen de entrada se introduce la imagen fusionada con el mapa de calor y tras estas 10 determinaciones se realiza la media de todos los resultados obtenidos en la clasificación celular como imagen "no revisable" (benigna) y "revisable" (ASC-US, LSIL y HSIL).

Para valorar los resultados obtenidos, se estudian las siguientes métricas: *exactitud, precisión (o valor predictivo positivo), exhaustividad (o sensibilidad) y puntuación F1*. Los resultados de la media de las determinaciones se muestran junto a su desviación estándar en la **Figura 5.10**. Se detalla una figura para cada métrica analizada, en la que se marca para cada valor de α la media de las 10 determinaciones realizadas, que se marca mediante un punto en el medio de una línea vertical que indica su desviación estándar.

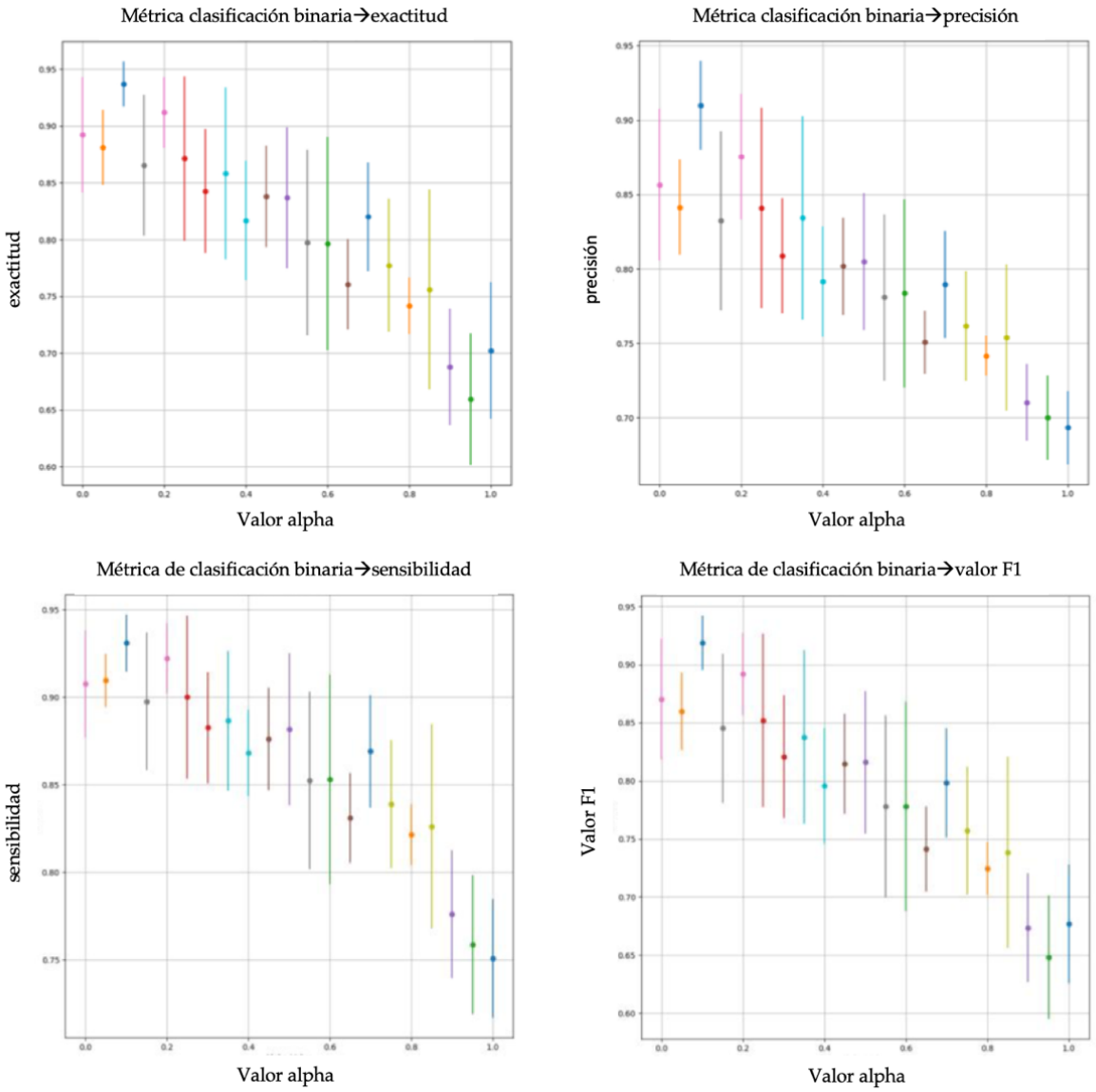


Figura 5.10. Métricas de los resultados obtenidos para la exactitud (arriba a la izquierda), precisión (arriba a la derecha), exhaustividad (abajo a la izquierda) y parámetro F1 (abajo a la derecha). En el eje Y se observa el valor de la métrica a estudio y en el eje X el valor del parámetro α. Cada línea de color determina el valor del parámetro a estudio para un determinado valor α, que presentan un punto central que señala la media de 10 determinaciones realizadas para cada valor de α. Elaboración propia.

De manera global se puede apreciar en la gráfica que conforme aumentamos el valor del parámetro α disminuye el valor de la métrica estudiada empeorando los resultados, debido a que el mapa de calor no proporciona suficiente información que cubra las pérdidas de los detalles valiosos que aporta la imagen original. Analizando en detalle, encontramos que el valor de transparencia óptimo en términos de clasificación es $\alpha = 0.1$. Con un valor de $\alpha = 0.1$ se obtiene una clasificación con una exactitud del 94% ± 2%, una precisión del 91% ± 3%, una exhaustividad del 93% ± 2% y una puntuación F1 del 92% ± 2%. Los resultados muestran que con este valor el número total de células bien clasificadas es de 113, y solo 7 están mal clasificadas.

En la prueba del modelo, tras la clasificación multiclase hemos simplificado los resultados en una clasificación binaria de células "revisables" y "no revisables". Para estudiar el verdadero beneficio que nos aportan las imágenes con su mapa de calor, en la **Tabla 5.13** planteamos una matriz de confusión que compara los resultados de las células sin mapa de calor ($\alpha = 0.0$) con las células fusionadas con el mapa de calor con el parámetro α óptimo ($\alpha = 0.1$). En la tabla se refleja la clasificación del modelo de las 120 células (30 de cada categoría diagnóstica); a la izquierda los datos obtenidos de la imagen original sin mapa de calor ($\alpha = 0.0$) y a la derecha los obtenidos de la imagen fusionada con $\alpha = 0.1$. Con un valor de $\alpha = 0.0$ se han clasificado correctamente 106 células de 120 (resultado del sumatorio de 28 células "revisables" y 78 células "no revisables" correctamente clasificadas) mientras que con un valor de $\alpha = 0.1$ se han clasificado correctamente 113 (resultado del sumatorio de 28 células "no revisables" y 85 "revisables"). Esta cantidad de células correctamente clasificadas representan el 88,3% y el 94,2% de la exactitud para $\alpha = 0.0$ y $\alpha = 0.1$, respectivamente (**Tabla 5.13 y 5.14**).

Tabla 5.13. Matriz de confusión que compara los resultados obtenidos, empleando imágenes con $\alpha = 0.0$ y $\alpha = 0.1$. Se detalla la exactitud o precisión global para cada valor del parámetro α. Elaboración propia.

REAL	**$\alpha = 0.0$** Exactitud=88%		**$\alpha = 0.1$** Exactitud=94%	
	No revisable	Revisable	No revisable	Revisable
No revisable	28	2	28	2
Revisable	12	78	5	85

La **Tabla 5.14** desarrolla el valor de las distintas métricas estudiadas, enfrentando los resultados obtenidos para $\alpha = 0.0$ y $\alpha = 0.1$. Destaca la precisión del algoritmo con el 97,5% y 97,7% para $\alpha = 0.0$ y $\alpha = 0.1$, respectivamente y una mínima tasa de falsos negativos para $\alpha = 0.1$ del 5.56%. En la **Tabla 5.15** se detalla el porcentaje de células bien y mal clasificadas para ambos valores del parámetro α, siendo del 88% para $\alpha = 0.0$ y 94% para $\alpha = 0.1$.

Tabla 5.14. Tabla con los parámetros obtenidos a través de la matriz de confusión de los resultados obtenidos para $\alpha = 0.0$ y $\alpha = 0.1$. Elaboración propia.

Métricas	**$\alpha = 0.0$**	**$\alpha = 0.1$**
VP	78	85
FP	2	2
FN	12	5
VN	28	28
Sensibilidad	86,67 %	94,44 %
Especificidad	93,33 %	93,33 %
Precisión	97,50 %	97,70 %
Valor predictivo negativo	70 %	84,85 %
Exactitud	88,33 %	94,17 %
F1	91,88 %	96,05 %
Tasa de falsos positivos	6,67 %	6,67 %
Tasa de falsos negativos	13,33 %	5,56 %

Tabla 5.15. Se detalla el porcentaje de células bien etiquetadas y mal etiquetadas en una clasificación binaria ("revisables" y "no revisables"), después del uso de imágenes originales (α = 0.0) y las imágenes fusionadas con α =0.1. Elaboración propia.

	α = 0.0	**α =0.1**
Bien clasificadas	88,33%	94,16%
Mal clasificadas	11,66%	5,83%

5.5. EL USO DE LA IA COMO HERRAMIENTA DIAGNÓSTICA EN PAÍSES EN VÍAS DE DESARROLLO

En países en vías de desarrollo presentan grandes limitaciones para el acceso al sistema sanitario. La IA se presenta como una herramienta muy útil en estos escenarios en los que las pacientes debutan con frecuencia en estadios muy avanzados de la enfermad por la escasez de recursos socio-sanitarios. La IA podría suplir en cierto modo a patólogos expertos en territorios donde son muy escasos, permitiendo detectar a las pacientes en riesgo de desarrollar enfermedad, logrando un diagnóstico y tratamiento precoz.

La principal limitación que podemos encontrar, es que en estos países el acceso a sistemas de digitalización de portaobjetos es limitado y de alto costo, por lo que se ha de recurrir a dispositivos de bajo rendimiento que obtienen imágenes de células de calidad algo inferior. Además, la aplicación de redes CNN de diagnóstico celular en estos países supondría un gran gasto de recursos difícil de implementar en estos escenarios. Proponemos un modelo de bajo coste como vía de solución en entornos de recursos limitados, manteniendo un buen rendimiento y fiabilidad del modelo.

5.5.1. MANEJO DE LA IMAGEN OBTENIDA POR UNA VIDEO-CAPTURADORAS DE MICROSCOPIO

Proponemos el uso de métodos de captura de imagen como las cámaras video-capturadoras para microscopio, que están disponibles en el mercado a precios muy económicos y proporcionan imágenes de buena calidad **(Figura 5.11)**. Se instalan en un microscopio trinocular y a través de un cable USB y mediante un software gratuito obtenemos la imagen en tiempo real en el ordenador (**Figura 5.12**).

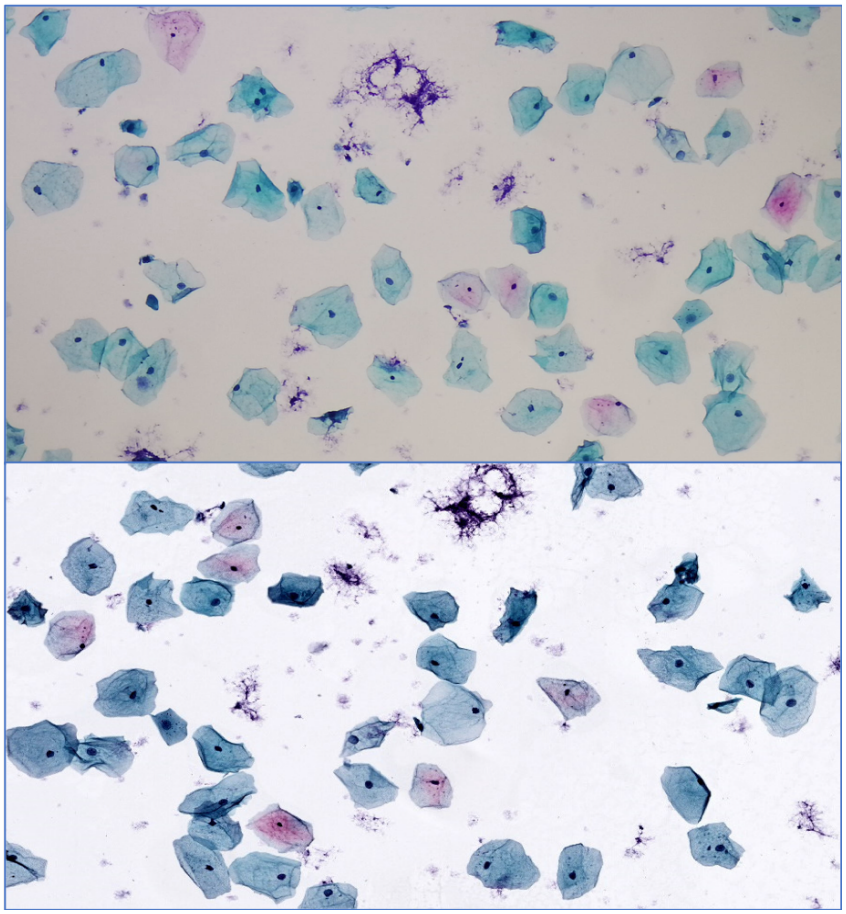


Figura 5.11. Comparativa de un campo digital de células obtenidas mediante una cámara low cost (arriba) frente a una laminilla escaneada con el escáner Ventana DP200 (abajo). Nótese la aceptable calidad de la imagen low-cost que; sin embargo, resulta mejorable al compararla con la imagen digital. Elaboración propia.

Debido a que las imágenes de las células son obtenidas por un dispositivo de peor resolución que el escáner de laminillas, se obtienen imágenes en buena calidad pero no óptima. Para optimizar la calidad obtenida hemos empleado *algoritmos de mejora de la imagen* que emplean redes antagónicas que compiten entre sí para obtener una imagen de calidad, que pueden usarse para simular un dispositivo de alto-coste (**Figura 5.13**), como veremos en el siguiente apartado.

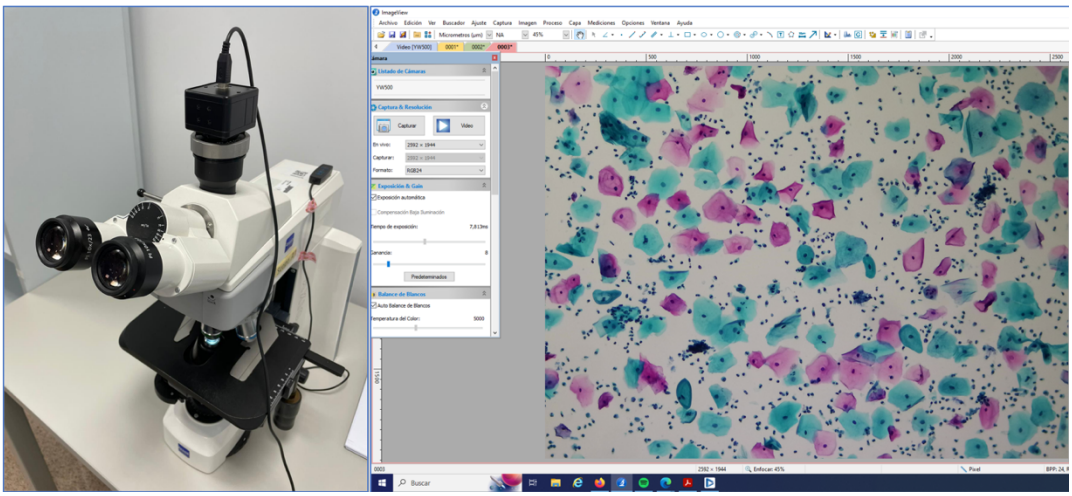


Figura 5.12. A la izquierda vemos un microscopio trinocular con una cámara video-capturadora low-cost instalada en su parte superior. A la derecha vemos el software gratuito que nos permite ver la imagen en tiempo real. Elaboración propia.

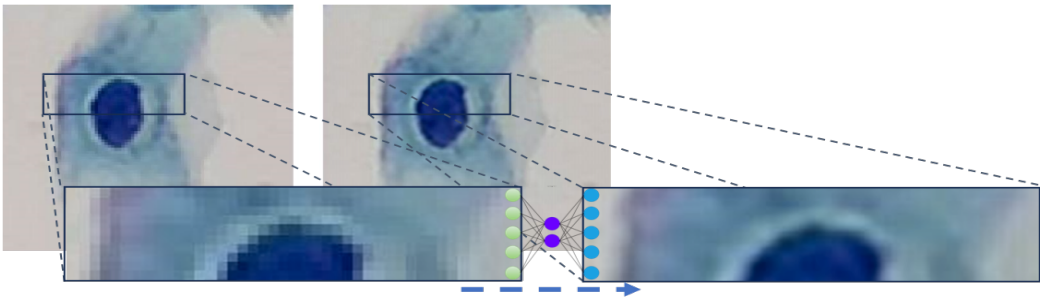


Figura 5.13. Algoritmo de mejora de la resolución de la imagen. A la izquierda una imagen del dispositivo low-cost original. A la derecha la misma imagen mejorada. Elaboración propia.

5.5.2. MANEJO DE LA IMAGEN OBTENIDA POR DISPOSITIVOS DE BAJO COSTE

Otro método de obtención de imágenes celulares en países en vías de desarrollo es a través de dispositivos de bajo coste como tabletas o teléfonos móviles con una calidad que; aunque es aceptable como en el caso anterior, es inferior a la calidad óptima.

Para solventar esta barrera, planteamos aplicar un algoritmo Real-ESRGAN para mejorar de la calidad de la imagen y testear a la red para comprobar si hay un cambio en la clasificación celular tras la mejora de la calidad de la imagen. La Real-ESRGAN (del inglés, *enhanced super-resolution generative adversarial network)* utiliza redes adversarias antagónicas que reconstruyen detalles poco nítidos compitiendo entre sí para logran imágenes realistas y más nítidas. Una de las redes genera detalles nítidos y la otra critica y valora si estas imágenes mejoradas parecen o no reales.

Para ello tomamos ejemplos de imágenes de mala, media y alta calidad y tratamos de mejorarlas, generando distintas bases de datos de calidades mejoradas y sin mejorar (**Figura 5.14**). También hemos probado a mezclar distintas calidades celulares (mala, media y alta); para posteriormente mejorarlas, obteniendo un grupo de células mezcladas; tanto mejoradas como sin mejorar (**Figura 5.15**). Estas bases de datos mixtas tratan de combinar distintas calidades para simular un entorno diagnóstico real, en el que dentro de una misma citología coexisten distintas calidades celulares.

Tras testear la red se ha evidenciado que el modelo más equilibrado es aquel que emplea una mezcla de imágenes de las tres calidades (baja, media y alta) una vez han sido mejoradas con Real-ESRGAN, con una precisión del 92,2%.

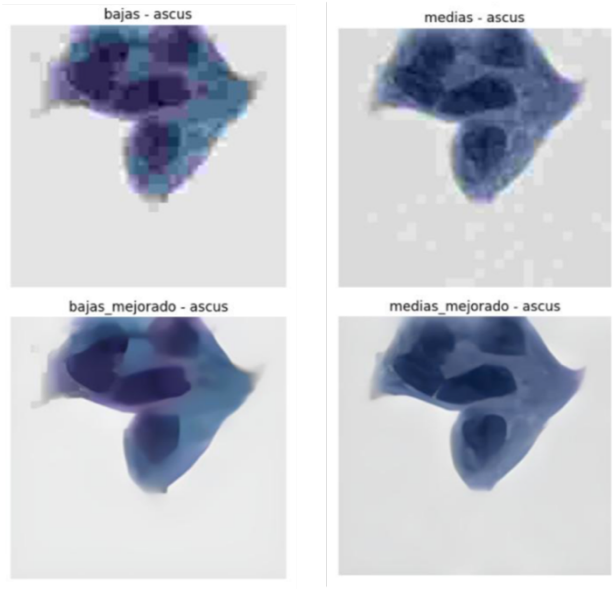


Figura 5.14. Imágenes de baja y media calidad de una célula ASC-US sin mejorar (arriba) y mejorada (abajo). Imagen obtenida del trabajo fin de grado "Sistema de diagnóstico para el cáncer de cérvix mediante dispositivos de bajo coste" de D. Juan Miguel Correa García. Elaboración propia.

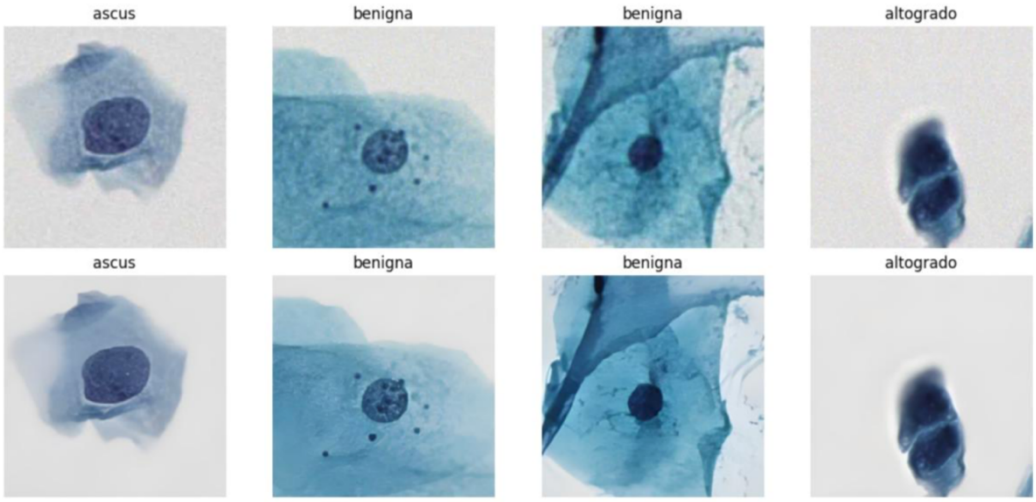


Figura 5.15. Imágenes de distintas calidades mezcladas sin mejorar (arriba) y mejoradas (debajo). Imagen obtenida del trabajo fin de grado "Sistema de diagnóstico para el cáncer de cérvix mediante dispositivos de bajo coste" de D. Juan Miguel Correa García. Elaboración propia.

Este hallazgo pone de manifiesto la importancia de trabajar en unas condiciones próximas a las reales, en las que aproximándonos lo máximo a la realidad, podemos obtener mejores resultados y conseguir modelos de fácil implantación en la clínica diaria.

5.5.3. DISEÑO DE UN MODELO DE BAJO COSTE

Presentamos un modelo DL que sea aplicable en un entorno real en países con recursos limitados, en donde se pueda emplear sobre dispositivos móviles económicos o video-capturadoras de microscopio, presentándose como una solución diagnóstica en estos países.

Los modelos de DL permiten identificar y clasificar anomalías celulares, pero presentan puntos de mejora como se ha desarrollado en el resto de este estudio como: escasez de células etiquetadas, interpretabilidad y explicabilidad de los diagnósticos. En nuestro nuevo modelo, combinamos el uso de redes generativas antagónicas y mapas de calor que le aportan explicabilidad al nuevo modelo. Estos modelos se retroalimentan y van aprendiendo, aumentando la precisión de la red y emplean métodos multitarea que permiten obtener un diagnóstico al mismo tiempo que se segmentan e interpretan las células.

Proponemos el uso de modelos computacionales de bajo-coste que permitan obtener buenos resultados a bajo costo. Para ello, diseñamos un modelo de bajo-coste que ha sido entrenado en la nube, usando servidores remotos con gran capacidad computacional ahorrando recursos, que hacen sus predicciones en dispositivos de bajo coste como un móvil o un ordenador (sin depender de la nube) (**Figura 5.16**).

Para ello, se entrena un *modelo de aprendizaje profundo pequeño o TinyML* que se trata de una tecnología que traslada los mecanismos de modelos ML a dispositivos de bajo consumo energético y recursos limitados. Se trata de una rama

de la IA que se encarga de ejecutar modelos ML en dispositivos de bajo coste y baja potencia computacional, como móviles o sensores. Esta tarea se consigue a través de la reducción de las necesidades y tamaño de la red usando recursos limitados, sin sacrificar en exceso su precisión.

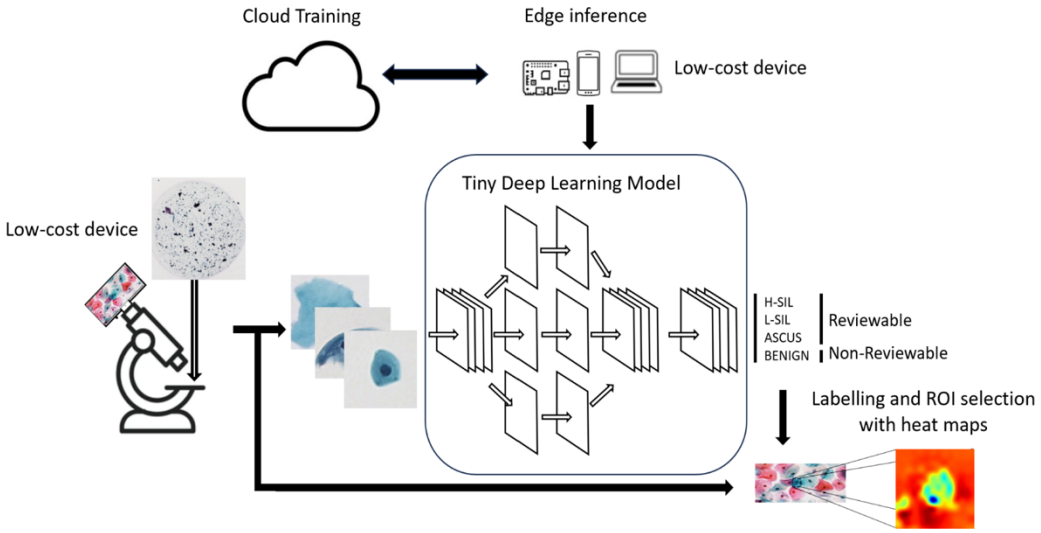


Figura 5.16. Modelo low-cost propuesto para el diagnóstico de citología cervicovaginal. Elaboración propia.

Concretamente, TinyML transforma modelos complejos DL como las redes CNN en versiones simplificadas de menor tamaño y complejidad, empleando menos memoria pero logrando un funcionamiento óptimo y reduciendo costes, objetivo fundamental para adaptar la CNN a dispositivos de bajo coste. A pesar de ser un modelo ligero y simplificado, sigue siendo capaz de etiquetar las células en su capa de salida como la red CNN de la que parte.

Una vez obtenemos nuestras imágenes, mejoramos su resolución mediante un algoritmo de mejora y las introducimos como imagen de entrada a la red TinyML, testeamos a la red para ver la precisión de sus resultados en la capa de salida. En su salida la célula será etiquetada en una de las categorías diagnósticas del Sistema Bethesda, que se reagrupará en un sistema de clasificación binario como "revisables" y "no revisables".

Para lograrlo, probamos distintas herramientas que permiten implementar la TinyML, para así poder aplicar modelos redes neuronales reducidas y optimizadas sobre dispositivos baratos de pequeño tamaño como móviles o cámaras sin conexión a internet. Estos modelos son: *TensorFlow lite* y *TensorFlow lite cuantizado*. Son modelos rápidos debido a que no necesitan enviar datos a la nube, ahorrando en datos. Se diferencian entre sí en que uno realiza una cuantización de los datos (convirtiendo números complejos con decimales en números pequeños enteros), simplificándolos aún más, acelerando la red y ocupando menos espacio y energía. A pesar de que el sistema de cuantización presenta el inconveniente de que puede perder ligeramente la precisión, esta pérdida no influye significativamente en el resultado final.

Se han probado diferentes métricas para estimar el mejor modelo de bajo coste: el modelo con *TensorFlow Lite* y el otro modelo más pequeño que incorpora la *cuantificación*. Estos modelos se han comparado con el *modelo original* en la **Figura 5.17.** El modelo *lite cuantificado* presenta el tamaño de archivo más pequeño, con una reducción de aproximadamente el 97% en comparación con el *modelo original*. Esta reducción es especialmente beneficiosa para entornos con almacenamiento limitado como los teléfonos móviles. Este modelo mejora la inferencia móvil en aproximadamente un 81% con respecto al *modelo lite*, lo que indica que la *cuantificación* es particularmente beneficiosa para el rendimiento móvil.

Ambos *modelos TensorFlow* optimizados utilizan significativamente menos RAM que el *original*, lo cual es crucial para ejecutar modelos en dispositivos con recursos limitados. La reducción de RAM en el *modelo lite cuantificado* es de aproximadamente el 60%, lo que lo hace especialmente adecuado para entornos con memoria limitada.

En términos de precisión, el *modelo original* tiene una precisión del 88% para la clasificación en cuatro clases (HSIL, LSIL, ASC-US y benigno) y del 97% para la clasificación binaria (que requieren o no revisión por parte del patólogo). Los *modelos lite* no muestran cambios significativos en la precisión, por lo que se ven afectados por una pérdida del 1 % y del 2 % con los *modelos lite* y *lite cuantificado*.

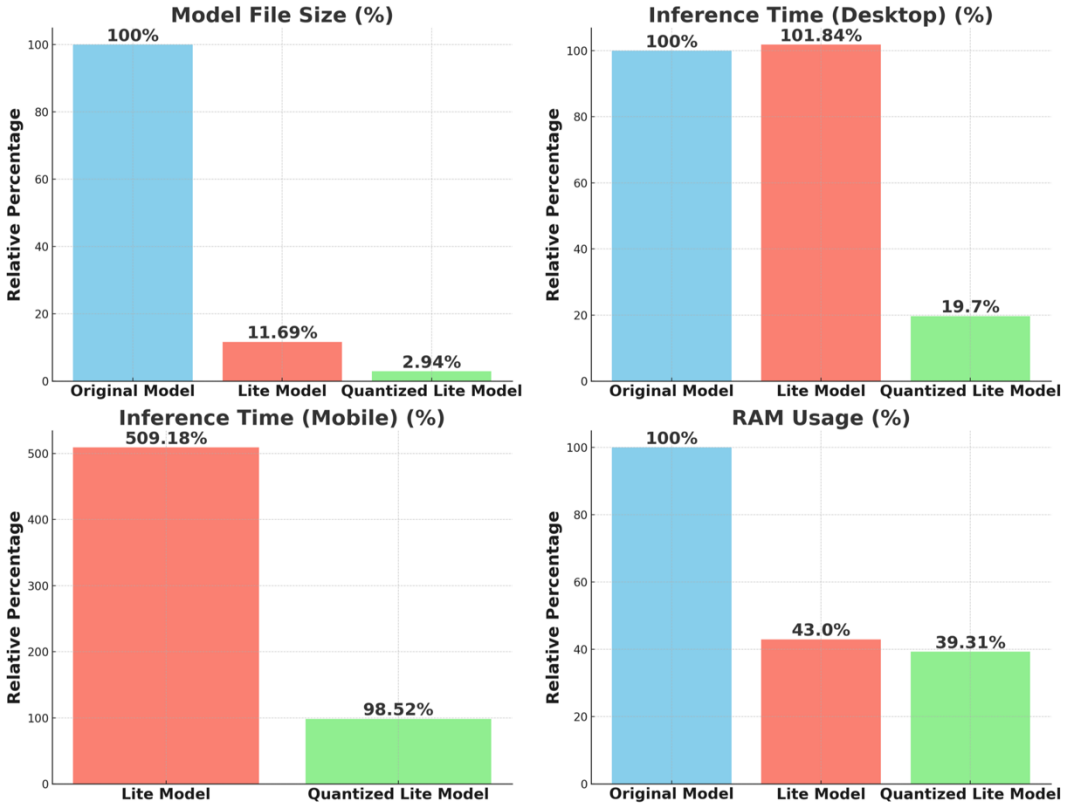


Figura 5.17. Resultados de comparar el modelo original (barra en azul), modelo TensorFlow lite (barra en rojo) y TensorFlow lite cuantizado (barra en verde). Elaboración propia.

Por otro lado, como en fases previas del estudio y para aportarle transparencia al modelo, planteamos la utilidad de la explicabilidad como herramienta para mejorar la confianza en el modelo. Mediante la fusión de la imagen de salida con mapas de calor la red nos marca en qué zonas de la imagen se está fijando para la toma de decisión, mejorando la confianza en el modelo (**Figura 5.16 y 5.18**). Como en otras fases del estudio, la red ha mostrado resultados excelentes en la creación de mapas de calor.

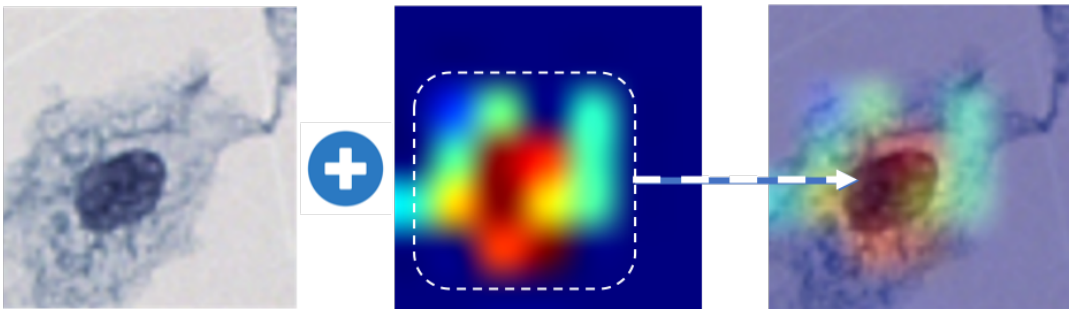


Figura 5.18. Modelo de explicabilidad mediante la fusión de la célula original y un mapa de calor. Elaboración propia.

VI – DISCUSIÓN

VI -DISCUSIÓN

6.1. EL USO DE IMÁGENES ORIGINALES NO PREPROCESADAS

Hay varios estudios publicados sobre el uso de algoritmos de diagnóstico celular que trabajan sobre conjuntos de datos de imágenes públicos (Taha et al., 2017) (X. Zhang et al., 2025) o de nueva creación que han sido preprocesados, obteniendo grandes resultados. Zhozaimi y cols. observaron una mejora de las métricas tras el uso de técnicas de mejora de la imagen del 13,62% en exactitud, 13,08% en sensibilidad y 14,34% en valor F1 (Khozaimi et al., 2025).

Hay varios trabajos que estudian los beneficios del preprocesamiento de imágenes en el diagnóstico celular. Sarhangi y cols. han enfrentado un conjunto de datos de imágenes preprocesadas SIPaK con otras imágenes sin modificación previa con unas precisiones del 99,6% frente al 96,5% respectivamente, lo que pone de manifiesto que; aunque el preprocesamiento se presenta como una técnica de mejora de la exactitud de la red, muestra una discreta mejora de los resultados a costa de una metodología de trabajo más compleja (Ahmadzadeh Sarhangi et al., 2024). Otro trabajo comparativo de la base de datos Herlev solo evidenció mejoras en las métricas tras el empleo de métodos de segmentación (Albzour & Lam, 2025). Del mismo modo, hay investigaciones que trabajan sobre laminillas completas escaneadas, como el sistema AICCS, obteniendo grandes resultados sin la necesidad del empleo de métodos de realce y mejora de la imagen (J. Wang et al., 2024).

En nuestro trabajo empleamos imágenes sin preprocesamiento previo, tan cual son obtenidas desde el escaneo del portaobjetos. Estas imágenes son fieles a las células reales obtenidas mediante el examen del cuello del útero y su tinción con la tinción de Papanicolau, con células plegadas, borrosas y/o superpuestas que dificultan la tarea de clasificación celular para el algoritmo.

A pesar de a priori dificultar la tarea diagnóstica al algoritmo, optamos por estas imágenes no preprocesadas porque son fieles a las células reales y una vez obtengamos un modelo robusto, son más fáciles de trasladar a la práctica diagnóstica diaria. De este modo, una vez que el algoritmo se enfrente a pacientes reales que acudan a realizarse su citología, los resultados serán fieles a los observados durante el entrenamiento de red.

6.2. MEJORA DE LA NUEVA BASE DE DATOS MEDIANTE LA CREACIÓN DE CÉLULAS ARTIFICIALES

La IA se plantea como una gran herramienta para la creación de imágenes artificiales. En el campo de la medicina presenta un valor sobreañadido al agilizar enormemente el flujo de trabajo gracias a la posibilidad de aumentar bases de datos en menor tiempo sin la necesidad de realizar trámites burocráticos para la recolección de células de pacientes reales.

Una citología contiene de media entre 50.000 y 200.000 células escamosas, de entre las cuales encontramos únicamente de 10-30 células ASC-US (del 0,02-0,05% de la celularidad total), de 100-300 LSIL (0,1-0.5% de la celularidad total) y de 100-200 células HSIL (0,05-0,3% del total de células). Estos datos ponen de manifiesto la dificultad en la recolección de células de cada categoría. En la categoría ASC-US, para obtener el sumatorio obtenido con la creación de células artificiales, si necesitáramos obtenerlo de células reales sería necesario el escaneo de hasta 200 muestras de pacientes consumiendo tiempo y recursos.

No se han encontrado artículos recientes sobre la creación de células artificiales. Los artículos más recientes se centran en desarrollar conjuntos de bases de datos con anotaciones como Cx22 Dataset, muy útiles para entrenar a los modelos de diagnóstico por imagen (J. Bai et al., 2025) (X. Zhang et al., 2025).

La creación de células artificiales se presenta como una solución prometedora en estudios en el que el acceso a imágenes reales es limitado, permitiendo equilibrar las bases de datos entre distintas patologías o categorías, para así obtener modelos robustos de clasificación que puedan ser implantados en la práctica diaria. En nuestro estudio; gracias a la creación de células artificiales, hemos mejorado nuestra base de datos disminuyendo los tiempos de trabajo requeridos para obtener tal cantidad de imágenes, eliminando problemas burocráticos de la obtención de imágenes médicas y obteniendo una base de datos equilibrada en la que cada categoría diagnóstica tiene el mismo peso en datos durante el entrenamiento.

Las células obtenidas son fieles a los criterios del Sistema Bethesda y muchas de ellas son indistinguibles de las células reales de la base previa del estudio por un patólogo experto. En las células artificiales vemos que es frecuente la presencia de superposición nuclear, agrupamientos celulares, plegamientos, elementos de fondo o defectos tintoriales del procesamiento, debido a que hemos entrenado las redes GAN con imágenes sin preprocesamiento previo.

6.3. POTENCIAL DE NUESTRO MODELO HÍBRIDO DCGAN EN EL DIAGNÓSTICO DE CÉLULAS CERVICALES

Nuestro modelo híbrido DCGAN ha puesto a prueba las células artificiales generadas para evaluar si son superponibles en calidad a las células reales. Para ello las hemos sometido a una clasificación al igual que estudios previos para ver cómo responde la red. Tras el entrenamiento de la red con células reales y someterla posteriormente a la clasificación de células artificiales, los resultados han sido prometedores, demostrando que la creación de células artificiales es una solución de calidad a la problemática de la recogida de imágenes de nuestra base de datos (Martínez-Más et al., 2020).

Hemos partido de una base de datos de 1405 células reales seleccionadas manualmente a una elaboración de 8000 células artificiales por la red GAN que; tras el descarte de aquellas células mal etiquetadas por la DCNN, ha resultado en una muestra final de 6089 células artificiales correctamente etiquetadas por el modelo híbrido. De este modo, hemos ganado 1773 células benignas artificiales (descartando 227), 1752 células artificiales HSIL (descartando 248), 1426 ASCUS (descartando 574) y 1138 LSIL (descartando 862) (**Tabla 5.1 y 5.2**). Estos datos revelan que la red GAN ha generado y etiquetado correctamente un 76% del total de las células elaboradas, que se tratan de células de gran calidad indistinguibles de células reales por patólogos expertos. Las redes GAN tienen la ventaja de que al competir entre sí, van aprendiendo y mejorando sus resultados conforme van elaborando células. Además, este sistema ML ofrece la ventaja añadida de poder trabajar sobre las imágenes originales sin preprocesamiento previo ahorrando tiempo y gasto computacional en etapas previas a la computación.

Los resultados (**Tabla 5.2** y **Figura 5.2**) muestran que el modelo es más preciso en la elaboración y etiquetado de las células benignas, seguidas por las HSIL. La red comete más errores en las categorías ASC-US y LSIL, siendo la categoría que peor genera y etiqueta la LSIL, que presenta un índice de aciertos del 56,9% y alcanza un índice total de fallos del 43,1% (**Tabla 5.5**). En la categoría LSIL, vemos que más de la mitad de los fallos corresponden con el etiquetado erróneo a ASC-US (585 casos que fueron etiquetados en el 29,25%, que representan el 67,8% del total de células erróneas). Respecto a los resultados observados para ASC-US, dentro de los errores, observamos que el fallo que más se repitió con gran diferencia fue el etiquetado erróneo a LSIL (427 casos que fueron etiquetados en el 21,35%, que representa el 74,39% del total de los errores) (**Tabla 5.5**).

Estos hallazgos son superponibles a los observados en la práctica diaria; siendo las categorías "intermedias" (ASC-US y LSIL) las que más discrepancia y variabilidad interobservador genera entre patólogos y menor confiabilidad diagnóstica, sobre todo ASC-US. Algunos estudios marcan el porcentaje de discordancia para el diagnóstico de ASC-US entre distintos observadores en el 50% (Alameda Q., F. y colaboradores, s. f.). Esto se debe a que los criterios citológicos

de la categoría ASC-US son difíciles de llevar a la práctica por ser sutiles y en ocasiones superponibles a las categorías benigna y LSIL; lo que lleva a las guías clínicas a recomendar el seguimiento de estas pacientes con nuevas citologías. En el caso de ASC-US gracias al seguimiento con citologías nuevas se demuestra con el tiempo que el 5-12% de las pacientes no tenían patología (benigna), que el 30-50% desarrollarán un LSIL y el 5-12% un HSIL, según datos de la AEPCC. La categoría LSIL representa a la lesión precursora del cáncer de cérvix en su estadio más inicial por lo que en ocasiones puede no estar representada completamente en la citología o aparecer estos cambios en muy escaso número de células. Esto explica por qué LSIL es la categoría con la que se parte con menor número de células reales al inicio del estudio con la digitalización de las muestras.

Otra de las posibles razones radica en que el término ASC-US es ligeramente subjetivo, en el que la presencia de atipia leve en las células escamosas que cualitativamente o cuantitativamente son insuficiente para catalogarlos como LSIL, es suficiente para su etiquetado como ASC-US. Es decir, pacientes con criterios sugerentes de LSIL pero que contienen muy pocas células problema o que se acompañan de otros factores que pueden favorecer las atipias reactivas como la presencia de infamación aguda, podrían ser categorizadas como ASC-US. La red en nuestro estudio no entiende de criterios de cuantificación ya que etiqueta cada célula de manera individual y obtiene buenos resultados. Por lo tanto, se espera que cuando la red incorpore criterios de cuantificación y valore la muestra en su totalidad en un entorno diagnóstico real, alcance resultados excelentes detectando prácticamente la totalidad de las pacientes con riesgo de desarrollar cáncer.

Es importante recordar que en nuestro caso trabajamos en el entrenamiento con imágenes en condiciones reales, en las que las células pueden acompañarse de otras células escamosas formando agregados, superponerse con células inflamatorias, hematíes, artefactos de laboratorio o mostrar plegamientos nucleares que dificultan el visionado de las células. Este condicionante se ve reflejado en el modelo, en el que también las células artificiales presentan dichos artefactos. El motivo es tratar de perfeccionar el modelo para acercarnos a un sistema automatizado de diagnóstico que se base en condiciones reales sin sesgos.

En escenarios de muestras de difícil visionado con imágenes borrosas, amontonamiento o superposición nuclear, que se tratan de condiciones reales frecuentes de diagnóstico al microscopio, el patólogo puede a través de la rueda micrométrica modificar el plano de enfoque de la célula problema dividiendo la imagen distorsionada de células superpuestas, en distintos planos celulares que permiten estudiar cada plano con más detenimiento facilitando el diagnóstico. Sin embargo, la red no puede realizar dicha tarea ya que dispone únicamente de una célula cuyo plano no puede modificar, con el riesgo de emitir diagnósticos erróneos (**Figura 6.1**). Este problema se ve reflejado en casos en que células benignas son clasificadas como HSIL debido a la superposición y amontonamiento, que dan falsas imágenes de hipercromasia y atipia nuclear (143 casos, que representan el 7,1% del total de las células benignas). Esta problemática también se da en el resto de categorías, en las vemos un etiquetado erróneo como HSIL en el 8,65% de las células LSIL (173 casos) y el 5% de las ASC-US (102 casos) (**Figura 6.1 y Tabla 5.5**).

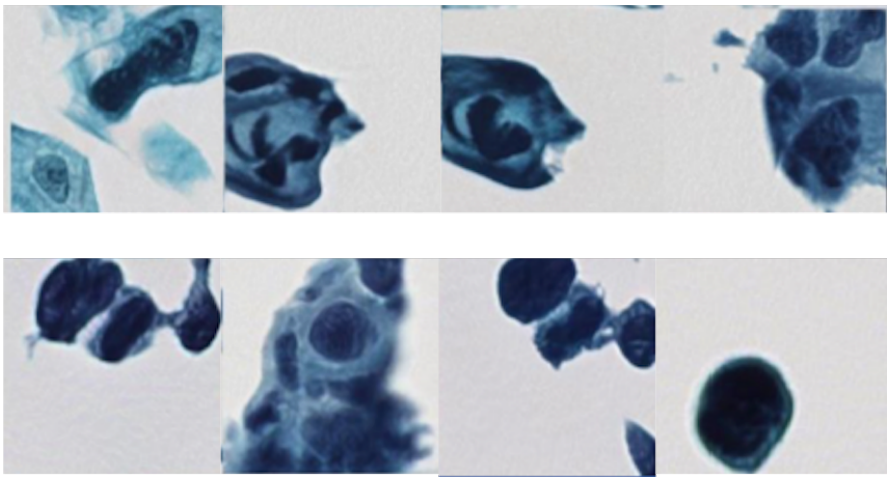


Figura 6.1. Imágenes de células artificiales cuya superposición celular, falta de nitidez o exceso de tinción podría justificar un etiquetado erróneo a HSIL, siendo originalmente creadas como LSIL (fila de arriba) y ASC-US (fila de abajo). Elaboración propia.

El análisis de métricas de la **Tabla 5.4** nos permite estudiar la fiabilidad del modelo para el diagnóstico de cada categoría diagnóstica de manera independiente:

- La clase benigna presenta las cifras más altas de sensibilidad (88,65%), especificidad (91,34%) y precisión (91,34%), lo que se traduce en que el modelo muestra una gran habilidad para detectar los casos benignos, cometiendo muy pocos falsos positivos, con una TFP baja, del 2.80%. Presenta un VPN del 96,25%, que revela que ante una predicción de la red de que la célula es benigna es muy probable que esto sea así, pudiendo descartar con seguridad la patología. Además, presenta un valor F1 del 89,98%, siendo el más alto de todas las categorías, lo que refleja un buen equilibro entre la precisión y sensibilidad en el que detecta casi todas las células que realmente son benignas.

- La clase HSIL; a pesar de tener unas cifras altas de sensibilidad (87,6%) y VPN (95,75%), presenta una cifra de precisión inferior a la categoría benigna (80,74%); que se traduce en que en algunos casos la red predice una célula como HSIL y finalmente no lo es (418 casos de 2000 células artificiales HSIL, lo que representa una TFP del 6,97%) (**Tabla 5.3**). Una de las razones que pueden provocar este fallo es (como se ha desarrollado anteriormente) que la presencia de hipercromasia y superposición nuclear puede dar falsas imágenes de atipia (**Figura 6.1**). A pesar de esta cifra de TFP, el valor F1 sigue siendo elevado, mostrando un buen equilibrio entre sensibilidad y precisión (84,03%).

- La clase ASC-US presenta unos resultados intermedios con cifras de sensibilidad y precisión moderadas (71,30% y 64,35%, respectivamente). Cifras como una TFP de 13.17% o TFN de 28,70% refleja que el modelo presenta dificultades en identificar esta categoría, de manera superponible a lo que pasa en ocasiones en la práctica diaria en la que la ambigüedad de esta categoría dificulta su clasificación. Del mismo modo el valor F1 del 67,65% confirma que no obtiene un equilibrio óptimo en esta clase.

- En el caso de las LSIL, se observa el peor rendimiento de todas las clases con una sensibilidad del 56,90%, precisión del 68,02% y valor F1 del 61,67%. Casi la mitad de los casos LSIL no son detectados, con una TFN del 43,10%. El VPN también muestra una cifra limitada del 86,38%.

De forma global, las métricas de la **Tabla 5.4** permiten concluir que el modelo es muy sólido para detectar los casos benignos y HSIL, sirviendo como un método fiable para "descartar" a las mujeres sanas y detectar a las pacientes con más riesgo de desarrollo de neoplasia, como las HSIL. Sin embargo y tal y como se ha discutido anteriormente en este apartado, la detección de ASC-US y LSIL supone una debilidad para la red, resultado de la similitud entre ambas categorías diagnósticas. Gracias a este método de desarrollo de imágenes artificiales y tras el descarte de aquellas imágenes erróneamente clasificadas, podemos emplear las células artificiales de las categorías más vulnerables para hacer un entrenamiento de calidad del modelo y equilibrar la muestra de células de entrenamiento entre cada categoría diagnóstica.

Para analizar la repercusión asistencial de la aplicación de nuestro modelo híbrido en un programa de cribado, podemos dividir las muestras en "revisables" y "no revisables". El programa de cribado del cáncer de cérvix pretende detectar a las mujeres con riesgo potencial de desarrollar cáncer. Las muestras "revisables" son aquellas que contienen cierto grado de displasia por lo que requieren seguimientos periódicos estrechos (ASC-US, LSIL y HSIL) frente a las "no revisables" que se corresponden a las citologías benignas de mujeres sanas que no necesitan una terapéutica adicional.

De manera práctica sería un problema que pacientes con citologías displásicas "revisables" fueran clasificadas erróneamente como benignas o "no revisables" ya que no seguirían controles a pesar de estar indicados y podrían desarrollar cáncer de cérvix sin ser detectadas. Según los datos obtenidos y si los distribuimos en una clasificación binara, han sido clasificadas erróneamente como "no revisables" o benignas: un 0,95% de las H-SIL (19 casos), un 2,25% para ASC-US (45 casos) y un 5,2% para L-SIL (104 casos) (**Tabla 5.2**). Corresponde a un global de 168 pacientes de 6000 "revisables" (2000 por cada una de las 3 categorías diagnósticas) que corresponde con un error de etiquetado del 2,8% del total de las células de estas 3 categorías (**Tabla 5.3** y **Figura 5.2**).

El caso a la inversa se trataría de pacientes con citologías "no revisables" benignas, que fueran etiquetadas como "revisables"; que correspondería con pacientes sanas que se harían seguimientos innecesarios no indicados por las guías clínicas con el gasto socio-sanitario que acarrea. Esta problemática es ligeramente menor al escenario previo en términos de morbilidad porque; al contrario que en el caso anterior, son mujeres sanas y la única consecuencia es la creación de "falsas enfermas" con la consecuente preocupación generada en las pacientes. Los resultados muestran que el etiquetado erróneo de pacientes "no revisables" a "revisables" es: 7% a H-SIL (143 casos), un 4% a ASC-US (78 casos) y un 0,3% a L-SIL (6 casos). El sumatorio de pacientes con muestras benignas que serían etiquetadas como "revisables" serían 227 pacientes, con un índice de error del etiquetado del 11,3% en este grupo (**Tabla 5.6** y **5.7**).

6.4. SEGMENTACIÓN COMO MÉTODO DE INTERPRETABILIDAD

La segmentación es una herramienta que nos permite indicarle a la red donde debe de fijarse para emitir el diagnóstico; es decir, nos permite enriquecer la imagen de entrada para mejorar los resultados en la salida.

Si comparamos la segmentación manual por un patólogo experto de la **Figura 5.3** y la segmentación automática por la red de la **Figura 5.4** vemos como las regiones marcadas por la red son superponibles a las marcadas por un patólogo experto y que la segmentación automática es totalmente equiparable a la manual, mostrando grandes resultados únicamente con un entrenamiento de 80 células (un 5% de nuestra base de datos). Esta alta precisión nos ha permitido combinar el uso de células segmentadas con las imágenes originales para crear un sistema de clasificación celular mejorado con un procesamiento de datos más fluido para así obtener un modelo más robusto.

A la hora de la clasificación celular tras la etapa de segmentación, hemos comparado los 3 modelos propuestos analizados desde dos métodos de clasificación celular: la clasificación multiclase y la binaria.

A. CLASIFICACIÓN MULTICLASE

Respecto al método de *clasificación multiclase*, hemos sometido los modelos al test de Wilcoxon para compararlos entre sí, del que extraemos las siguientes conclusiones:

- Al comparar el modelo clásico con el modelo 1, se obtiene una p = 0.1. Significa que hay evidencia moderada de que el *modelo clásico* es peor que el *modelo 1* con un nivel de confianza del 90%, no siendo una diferencia estadísticamente significativa.

- Al comparar el *modelo clásico* con el *modelo 2*, se obtiene una p=0,05. Implica que el *modelo clásico* es significativamente peor que el *modelo 2,* con un nivel de confianza del 95%.

- Al comparar el *modelo 1 y 2* no existen diferencias estadísticamente significativas.

De los resultados, extraemos que los resultados de los *modelos 1 y 2* son superiores a los del *modelo clásico* con diferencias estadísticamente significativas en el caso del *modelo 2* tras su análisis mediante el test de Wilcoxon. Cuando comparamos los modelos 1 y 2 entre sí, no se observan diferencias estadísticamente significativas. Sin embargo, el *modelo 2* es más adecuado ya que presenta métricas ligeramente superiores a las del *modelo 1*, y también es computacionalmente más eficiente gracias a su capa GAP. Además, se observan diferencias en términos de

tiempo computacional, tardando el *modelo cásico* 0,083 segundos en inferir una imagen, 0,1447 segundos el *modelo 1* y 0,1241 segundos el *modelo 2*.

Las **Figuras 5.6** y **Tabla 5.8.** hacen un análisis de métricas de la clasificación multiclase. En todas las métricas, los modelos propuestos (*1 y 2*) superan al método cásico. Concretamente, el *modelo 2* consiguió las puntuaciones más altas, con una exactitud de 0,725 y una puntuación de F1 de 0,727. El *modelo 1* aunque presenta unos resultados inferiores al *modelo 2*, logra unos resultados satisfactorios con 0,70 de exactitud y 0,702 de la puntuación F1. Las cifras reflejan que los *modelos 1 y 2* mejoran los resultados de la clasificación multiclase, particularmente en la distinción entre categorías complejas con superposición de características celulares, al compararlo con el *modelo clásico*. El *modelo clásico* presenta únicamente una exactitud de 0,675 y una puntuación F1 0,667, con una precisión y sensibilidad sensiblemente menores, lo que podría conducir a un mayor número de falsos positivos o negativos en los casos más ambiguos.

B. CLASIFICACIÓN BINARIA

Cuando comparamos los resultados obtenidos en la clasificación binaria entre los modelos propuestos a través de la prueba de Wilcoxon, encontramos que:

- Al comparar el *modelo clásico* con el *modelo 1* obtenemos una p=0,03. Significa que el *modelo clásico* es significativamente peor que el *modelo 1* con un nivel de confianza del 97%.

- Al comparar el *modelo clásico* con el *modelo 2* obtenemos una p=0,01. Implica que el *modelo clásico* es significativamente peor que el *modelo 2*, con un nivel de confianza del 99%.

- De nuevo, al comparar el *modelo 1 y 2* en la clasificación binaria no existen diferencias estadísticamente significativas.

En este caso, los *modelos 1 y 2* siguen siendo superiores al *modelo clásico* mostrando en ambos casos diferencias estadísticamente significativas (p=0,03 y p=0,01, respectivamente). Al enfrentar los *modelos 1 y 2* con los resultados, el *modelo 2* destaca con cifras impresionantes, aunque no existe una diferencia estadísticamente significativa al compararlo con el *modelo 1* mediante el test de Wilcoxon; tal y como ocurre en la clasificación multiclase. Este análisis subraya la robustez de los modelos propuestos que utilizan la segmentación en el proceso de entrenamiento de la clasificación binaria, en comparación al *modelo clásico*.

En el caso de la clasificación binaria, atendiendo a las métricas, vemos en los resultados reflejados en la **Figura 5.7** y **Tabla 5.8** que las diferencias entre el *modelo clásico* y los modelos propuestos sigue siendo notoria, con una gran precisión y rendimiento de los *modelos 1 y 2*. El *modelo 2* presenta una gran precisión de 0,942 y una puntuación F1 de 0,943 lo que sugiere una alta fiabilidad en tareas de clasificación binaria donde la diferenciación precisa es crucial. El *modelo 1* también muestra un gran rendimiento con una precisión de 0,933 y una puntuación F1 de 0,935, lo cual lo hace competitivo. En el caso del *modelo clásico,* si bien es aceptable, con una precisión de 0,850 y una puntuación F1 de 0,858, se queda atrás de los modelos propuestos. Los resultados respaldan que los *modelos 1 y 2* contribuyen al equilibrio y mejora de las métricas estudiadas gracias a la aportación de la segmentación en el entrenamiento en la clasificación binaria, que le da solidez a los modelos propuestos y reducen las clasificaciones celulares erróneas.

La **Tabla 5.8** que compara las métricas de los modelos enfrentando la clasificación binaria y multiclase, demuestra que se produce una mejora de las métricas al pasar de la clasificación multiclase a la binaria en todos los modelos testeados:

- La AUC oscila entre 0,866 y 0,919 (*modelos clásico y 2*, respectivamente) y asciende a 0,939 en el *modelo 2* en la clasificación binaria; reflejando una mejora en la capacidad discriminatoria entre clases.

- El valor F1 asciende de 0,666 en la clasificación multiclase (*modelo clásico*) a 0,858 en la binaria (*modelo clásico*); alcanzando cifras muy altas, como 0,943 en el *modelo 2*.

- Respecto a la sensibilidad (o *recall*), se produce una mejora sustancial en los 3 modelos al pasar del método multiclase al binario (con ascensos de 0,675 a 0,850 para el *modelo clásico*; paso de 0,708 a 0,933 en el *modelo 1* y de 0,725 a 0,942 en el *modelo 2*). Esta mejora en las cifras de sensibilidad se traduce en una mejora de la capacidad de detectar los casos positivos, reduciendo la TFN, característica fundamental de una técnica de cribado de calidad.

- En el caso de la precisión también se produce una mejora significativa de las métricas con un máximo de 0,747 en la clasificación multiclase y de 0,945 en la binaria para el *modelo 2*. La posible razón de la mejora de dichos resultados es que al pertenecer las categorías más vulnerables de la clasificación multiclase (ASC-US y LSIL) a la misma categoría ("revisable") junto a una categoría con grandes resultados como HSIL, los fallos se equilibran y la tasa de error disminuye.

- Por último, la métrica *accurancy* (exactitud), refleja una mejora significativa en todas las determinaciones (con una mejora promedio de 0,18). Se observa una gran mejora en el *modelo 2*, que alcanza 0,942 y refleja una tasa de diagnósticos correctos muy alta.

Como interpretación global en el análisis de la clasificación multiclase y binaria observamos claras ventajas de la clasificación binaria: una mejor discriminación, una mejora del valor F1 (con mejor equilibrio entre precisión y sensibilidad) y menores errores de clasificación. También se reduce la ambigüedad de diagnósticos intermedios como ASC-US y LSIL que; al englobarlos en una única categoría "revisable", mejora la precisión del modelo.

Si analizamos el impacto de la segmentación en los resultados de la red tras la evaluación de los distintos modelos propuestos, observamos que al segmentar con precisión las regiones relevantes de la imagen, la U-Net influye directamente en el rendimiento del modelo de clasificación posterior (*modelos 1 y 2*). Una segmentación eficaz permite al sistema centrarse en las estructuras más informativas, reduciendo el impacto del ruido y mejorando la extracción de características. De este modo, la U-Net actúa como un filtro que facilita al clasificador el análisis de patrones visuales clave para la toma de decisiones, lo que resulta en una mayor precisión. Cuando no se utiliza la segmentación proporcionada por la U-Net (*modelo clásico*), se observa que el clasificador se ve obligado a interpretar las imágenes sin la ayuda de la morfología presentada, lo que dificulta su capacidad para identificar correctamente las clases.

Los resultados demuestran que la inclusión de la U-Net en el diagnóstico aporta un valor añadido al proceso de clasificación. Las técnicas de interpretabilidad en los modelos de segmentación no solo ayudan a explicar los resultados obtenidos, sino que también facilitan la validación clínica y fomentan la confianza entre profesionales médicos en el uso de estas herramientas de IA en el diagnóstico por imagen. Al pasar de una caja negra a un modelo con transparencia, permite a los médicos visualizar las áreas clave en las que el modelo basó su decisión, mejorando la confianza en mismo.

Hay publicados algunos estudios que emplean la segmentación celular como método de interpretabilidad en el diagnóstico celular que se comparan con resultados obtenidos por métodos de diagnóstico DL tradicionales. Según lo publicado por Albzour y cols., en su estudio tras el empleo de la segmentación en una clasificación binaria se produce un empeoramiento de los resultados con una exactitud del 81,02% en la base de datos no segmentada y del 80,80% en la segmentada (Albzour & Lam, 2025); frente a nuestros resultados con una exactitud del 85% en nuestro modelo clásico frente al 94,2% en nuestro *modelo 2* que incorpora la segmentación (**Tabla 5.8**). En nuestro estudio se demuestra un porcentaje de mejora cercano al 10% tras incorporar la segmentación (Ortiz-González et al., 2025).

C. COMPARATIVA DE LA CLASIFICACIÓN DEL MEJOR MODELO PROPUESTO

Una vez analizados los modelos, hemos continuado el estudio con el mejor de ellos; *el modelo 2*, y hemos estudiado las diferencias en sus resultados cuando se realiza una clasificación multiclase o binaria. En la multiclase hemos empleado 30 células por cada categoría diagnóstica (en total 120 células) y en la binaria 90 células "revisables" y 30 "no revisables" (en total otras 120 células). Los resultados se presentan en la **Figura 5.8**.

La matriz multiclase (**Figura 5.8-b**) muestra que; aunque el *modelo 2* es eficaz en rasgos generales, presenta dificultades para distinguir algunas de las categorías diagnósticas entre sí:

- Funciona bien con las células benignas con un gran rendimiento y un alto índice de verdaderos positivos, por lo que es capaz de seleccionar aquellas pacientes sanas, libres de riesgo de presentar displasia y que no requieren una terapéutica o seguimiento específico, reduciendo el riesgo de intervenciones innecesarias.

- Los HSIL contienen una tasa de verdaderos positivos importante, ya que son aquellas que tienen una mayor relación núcleo-citoplasma y alteraciones importantes en el contorno nuclear, que se ponen fácilmente de manifiesto mediante la segmentación.

- La categoría ASC-US, a menudo se caracteriza por un ambigüedad diagnóstica, clasificándose erróneamente con frecuencia como LSIL (9 casos) o benigno (4 casos), que podría ser preocupante desde el punto de vista clínico, ya que las pacientes benignas se realizan revisión a los 3 años, frente a las ASC-US y LSIL que requieren una nueva citología a los 6 meses. Por lo tanto, existe el riesgo de que se pasen por alto casos que podrían requerir un seguimiento adicional.

- En el caso de las LSIL se clasifican con frecuencia como ASC-US (6 casos), que afortunadamente llevan el mismo seguimiento clínico, un control citológico a los 6 meses.

Mediante la matriz de confusión hemos elaborado una tabla con las métricas más importantes del *modelo 2* multiclase que se exponen en la **Tabla 5.9** y **Figura 5.6:**

- La categoría benigna destaca por un índice alto de VP (28) y escasos FN (2). Además, presenta en general las mejores métricas de las 4 categorías, con una sensibilidad del 93,33%, especificidad del 94,44% y VPN del 97,70%; lo que refleja la gran capacidad de la red de detectar las células benignas y descartar todas aquellas que no lo son. Se observa un ligero descenso de la precisión (84,85%) y el valor F1 (88,89%) debido a una pequeña proporción de casos detectados como benignos que realmente pertenecen a otra categoría; es decir por los 5 FP que presenta (en este caso 4 casos como ASC-US y 1 como LSIL).

- La categoría HSIL presenta cifras excelentes de VN (89 casos) y una especificidad muy alta (98,89%), que refleja la gran capacidad de la red de detectar los casos que no son HSIL; es decir, no confunde otros tipos de células como HSIL. La sensibilidad más baja (66,67%) y alta precisión (95,24%) dejan ver que; aunque la red no tiene una gran habilidad en detectar los HSIL, cuando emite este diagnóstico es muy probable que sea correcto.

- LSIL contiene 22 VP y 8 FN, que refleja que aunque la red identifica bien este tipo celular aún tiene limitaciones, con una sensibilidad limitada al 73,33%. Por el contrario la especificidad es algo mayor (82,22%), por lo que la red tiene más destreza en detectar las células que no pertenecen a esta categoría. La precisión y valor F1 relativamente bajos (57,89% y 64,71%, respectivamente) sugieren que parte de los diagnósticos LSIL realmente no lo son (16 FP, repartidos entre HSIL y ASC-US).

- Respecto a ASC-US los resultados son algo inferiores con 17 VP, 13 FN y una sensibilidad del 56,67% que demuestra las limitaciones de la red con esta categoría. Sin embargo, presenta una buena capacidad para detectar las células que no son ASC-US (especificidad 87,78%). Observamos 11 FP, una precisión del 60,71% y un valor F1 del 58,62%, que muestra que la red falla etiquetando células como ASC-US que realmente no lo son.

En general la matriz multiclase y sus métricas sugieren que, si bien el modelo puede diferenciar ampliamente las categorías benignas de las HSIL, los límites de las categorías intermedias están poco claros y son más complejos, por lo que se requerirá más datos para mejorar su precisión. De este modo, los subtipos diagnósticos que generan más confusión para la red son ASC-US y LSIL, debido a que tienen en ocasiones superposición de las características nucleares y citoplasmáticas que hacen difícil su distinción, incluso para patólogos expertos. Factores como baja celularidad o mala calidad de la célula problema (con áreas de plegamiento, artefactos, superposición nuclear o falta de contraste de la imagen) pueden dificultar mucho la visualización y diagnóstico celular. Al trabajar con imágenes no preprocesadas nos arriesgamos a enfrentarnos a estas limitaciones. Sin embargo, consideramos que el uso de células no preprocesadas es el método de estudio más fiel a las condiciones celulares reales, por lo que sigue siendo nuestro método de trabajo de elección. Además, el volumen de datos para cada categoría diagnóstica es muy desigual, porque las células LSIL y especialmente ASC-US son muy infrecuentes en la muestra, representando solo alrededor de 5 células de aproximadamente 5000 que una citología puede contener.

Por el contrario, en la matriz de confusión binaria del *modelo 2* (**Figura 5.8-a**) que simplifica los resultados en "revisables" y "no revisables", vemos que presenta un alto rendimiento con unas cifras de clasificación errónea mínimas (**Tabla 5.10**).:

- La categoría ”revisable” muestra una alta sensibilidad (94,44%) para detectar aquellas pacientes que requerirán seguimiento estrecho (“revisable”), clasificando correctamente 85 pacientes sobre 90 (VP) frente a las 5 mal clasificadas (FN). Su alta especificidad (93,33%) se traduce en una gran capacidad para detectar los casos que no son revisables, con únicamente 2 FP. Además, tanto la precisión como el valor F1 se mantienen altos (97,70% y 96,05%, respectivamente) que indica que cuando la red marca un caso como “revisable” es muy probable que realmente lo sea.

- La categoría “no revisable” (equivalente a las células benignas de la clasificación multiclase) muestra una alta sensibilidad (93,33%) y especificidad (94,44%) con 28 clasificaciones correctas sobre 30. Se traduce en que la red es hábil en detectar los casos “no revisables” y los que no lo son. Únicamente 5 casos se etiquetaron como “no revisables” de manera errónea (FP) lo que representa un 6,67% de TFN. La red muestra una precisión y valor F1 ligeramente inferiores (84,85% y 88,89%, respectivamente) que se justifica que hay casos que etiqueta como “no revisables” que finalmente no lo son (FP).

Al comparar entre sí los dos sistemas de clasificación del *modelo 2* podemos destacar (**Tabla 5.8 y Tabla 5.11**):

- La categoría benigna o “no revisable”, presenta resultados equiparables en los dos sistemas de clasificación de resultados, por lo que; aún siendo una especificidad muy alta en ambos casos, no se observa un predominio de una clasificación sobre la otra.

- La categoría “revisable” que abarca ASC-US, LSIL y HSIL; presenta resultados superiores en la clasificación binaria, con una de aciertos de 85 casos (94%) frente a 59 casos de la clasificación multiclase (65%) tras el sumatorio de las cifras de ASC-US, LSIL y HSIL.

En el análisis de datos cuando enfrentamos ambas clasificaciones, hemos obtenido unos resultados más sólidos y fiables en la clasificación binaria, con un porcentaje de acierto al diagnóstico entre el 93% (células "no revisables") y el 94% (células "revisables") (**Tabla 5.11**). Por lo tanto; aunque la clasificación binaria puede parecer una clasificación demasiado simple al compararla con la multiclase con todas las consideraciones que hace el Sistema Bethesda, se trata de un sistema de clasificación muy práctico y funcional desde el punto de vista asistencial que muestra grandes resultados y es sencillo de aplicar a la práctica diaria.

Esta mejora de los resultados en la clasificación binaria es más notable en ASC-US y LSIL, ya que junto al HSIL están incluidos dentro de la misma categoría: "revisable". Esta clasificación simplificada es además una clasificación más práctica, ya que el objetivo del cribado del cáncer de cérvix es detectar a las pacientes con riesgo potencial de desarrollar cáncer, que quedan destacadas en la categoría "revisable" incluyendo ASC-US, LSIL y HSIL para realizar pruebas complementarias (como la determinación de HPV) y continuar con un seguimiento estrecho.

Por otro lado, la clasificación binaria al ser un modelo más simple, puede pasar por alto matices entre el resto de las categorías del grupo "revisables" (LSIL, ASC-US, HSIL) que; aunque todas ellas pertenecen al grupo revisable, pueden tener distinta implicación clínica, como el control con una nueva citología en los ASC-US frente al examen del cuello del útero y toma de biopsia mediante una colposcopia en HSIL. De esta forma a pesar de que el modelo binario es sólido y práctico, el modelo multiclase de 4 categorías da una perspectiva con más matices que todavía necesita un mayor refinamiento para lograr un nivel equiparable de fiabilidad.

Analizando desde el punto de vista clínico los posibles errores del etiquetado, concretamente la confusión entre ASC-US y LSIL; la Sociedad Española de Citología, considera en su Programa de Calidad que este tipo de error se trata de un *grado de confusión menor* debido a que presentan un seguimiento clínico similar: seguimiento con control citológico en 6 meses. Un factor que puede haber influido

en esta debilidad en la clasificación multiclase es la desigualdad entre el número de células de cada categoría durante el entrenamiento. Partíamos de una base de datos algo desequilibrada, de 389 HSIL, 293 ASC-US, 182 LSIL y 879 células benignas de las cuales se reservaron 20 células de cada grupo para la segmentación manual y entrenamiento de la red para lograr la segmentación automática. Una solución podría ser aumentar y equiparar la recopilación de células de las categorías minoritarias.

El grupo de trabajo de Austin y cols., utiliza una red CNN preentrenada mediante aprendizaje por transferencia para el diagnóstico celular de citologías cervicovaginales utilizando los conjuntos de datos Herlev y SLIPaKMeD, comúnmente utilizados en los proyectos de investigación de citologías cervicales. Además, compara los resultados del empleo de células segmentadas manualmente con células íntegras sin segmentar. En sus resultados de clasificación binaria (anormal vs normal) demuestra que no hay cambios significativos tras realizar la segmentación, con cifras de precisión en la base de datos SIPaKMeD para células segmentadas del 96,74%, frente al 97,55% de las no segmentadas (Austin & Parvathi, 2025). En nuestro proyecto; no solo hemos testeado la influencia de la segmentación en el diagnóstico, sino que hemos analizado el efecto de la segmentación automática (U-Net) en el mismo, comparando 3 modelos distintos entre sí. El *modelo 2* (que combina la U-Net con una capa GAP como clasificador final) ha obtenido excelentes resultados tras la segmentación, con una precisión del 97,70% en la clasificación binaria ("revisable" vs. "no revisable"), frente a un 88,9% de precisión empleando células no segmentadas (*modelo clásico*) (**Figura 5.7 y Tabla 5.8**).

6.5. LOS MAPAS DE CALOR COMO MÉTODO DE EXPLICABILIDAD

En la etapa de explicabilidad con la elaboración de los mapas de calor con Grad-CAM, la red ha marcado de forma visual y clara las zonas más relevantes de la célula para el modelo, tal y como haría un patólogo experto. La Grad-CAM nos permite evaluar cómo pequeñas modificaciones en la imagen producen una

modificación significativa a la salida del modelo, destacando zonas de la imagen cruciales para el diagnóstico como el núcleo.

En nuestros resultados, los mapas de calor generados muestran una correlación clara con los criterios empleados por un patólogo experto. Estos mapas de calor han sido validados por un patólogo experimentado, presentando una alta precisión y permitiendo su aplicabilidad en entornos clínicos y de docencia reales. Este método de explicabilidad; no solo destaca como herramienta diagnóstica real, sino que también deja ver cómo el algoritmo; sin tener conocimientos del campo médico, es capaz de apreciar la zona crucial de la imagen tal y como se describe en las guías de diagnóstico del Sistema Bethesda. Este hallazgo facilita la aceptación y la pérdida de desconfianza del personal sanitario al uso de estas redes en el diagnóstico por imagen ya que el modelo deja de funcionar como una caja negra, mostrando visualmente las regiones de la imagen que han impulsado su decisión diagnóstica.

En la etapa de clasificación se testea la utilidad de la elaboración de mapas de calor y tratamos de analizar cual es el mejor valor para el parámetro α que nos proporcione los mejores resultados en la clasificación celular aprovechando la información que nos proporciona tanto el mapa de calor como la imagen original (**Figura 5.9**). Los resultados ponen de manifiesto el impacto positivo que genera enriquecer las imágenes celulares con mapas de calor, especialmente cuando se emplean valores para α bajos.

En general, cuando empleamos un parámetro α demasiado bajo, el mapa de calor se centra más en el citoplasma y cuando se usan mayores valores para α, el mapa de calor le resta importancia al citoplasma y se centra casi exclusivamente en el núcleo, borrándose el resto de detalles de la célula, como la forma del núcleo o el contorno del citoplasma, lo que implica una pérdida importante de información de la imagen original. El valor óptimo de fusión es $\alpha = 0.1$ que mejora significativamente los resultados de clasificación presentándose como una herramienta útil para detectar los casos positivos y negativos, tal y como se refleja en las **Tablas 5.13** y **5.14**.

Si observamos la matriz de confusión de la **Tabla 5.13** vemos que en la clasificación de células "no revisables" (que se corresponde con las células benignas), se obtiene el mismo resultado independientemente del parámetro α, por lo que ambos parámetros α tienen una gran capacidad para detectar los casos negativos, con una especificidad del 93,3% (**Figura 5.9**). Donde se produce una diferencia notoria es en las células "revisables" en las que con $\alpha = 0.0$ un 13,33% de las células fueron mal clasificadas (12 células); mientras que con $\alpha = 0.1$ se obtuvo únicamente un 5,55% de error (5 células), suponiendo menos de la mitad de fallos (**Tabla 5.13**). Esta disminución de los errores al modificar el valor de α refleja la utilidad de los mapas de calor en el diagnóstico celular, siempre y cuando se mantenga un nivel de transparencia suficiente para no ocultar en exceso los detalles celulares de la imagen original.

La mejora en la sensibilidad entre el modelo de $\alpha = 0.0$ y $\alpha = 0.1$; con un aumento de 86,67% a 94,44% respectivamente (**Tabla 5.14**), presenta una gran relevancia a nivel clínico, ya que refleja una mejor capacidad de la red para detectar a las pacientes "revisables" que son las que precisan un seguimiento estrecho. La puntuación F1 también presenta una ligera mejora (de 91,88 a 96,05%, respectivamente), que refleja un buen equilibrio entre precisión y sensibilidad. De manera global, los resultados son superiores para $\alpha = 0.1$ a excepción de la especificidad, que en ambos casos muestra una cifra similar. Para $\alpha = 0.0$ la sensibilidad ha resultado ser la métrica con peores resultados (86,67%), ya que de los realmente positivos el modelo pasó por alto y no detectó 12 (FN); es decir, pasó por alto un 13,33% de las pacientes con citologías "revisables" que fueron diagnosticadas como "no revisables" (TFN). En la práctica clínica se traduce en pacientes que necesitan seguimientos periódicos que no se llevan a cabo, con el riesgo consiguiente de desarrollo de cáncer de cérvix sin posibilidad de detección temprana.

Los resultados ponen de manifiesto la mejora de la métricas ante el uso de las imágenes enriquecidas con el mapa de calor con valores de α bajos, respecto al entrenamiento único con imágenes originales (α de 0.0); con un índice global de células mal clasificadas (tras el sumatorio de los errores de clasificación de las

células "revisables" y "no revisables") del 5,83% para $\alpha = 0.1$ (7 células) frente al 11,66% para $\alpha = 0.0$ (14 células) **(Tabla 5.15).** De esta forma se demuestra que aplicando un valor óptimo de $\alpha = 0.1$, que emplea una influencia del mapa de calor del 10% sobre la imagen original, se obtienen resultados satisfactorios con un error de clasificación bajo. Estos resultados son prometedores y plantean la necesidad de explorar la posibilidad de ampliar la arquitectura a una clasificación de 4 clases.

Cuando se emplean valores de α demasiado altos (mayores de $\alpha = 0.1$), el mapa de calor entorpece el diagnóstico; sobre todo en diagnósticos en los que el aumento de la relación núcleo-citoplasma (que caracteriza a las células ASC-US, LSIL y HSIL) es crucial para el diagnóstico, ya que se borra la interfase del núcleo y citoplasma celular fusionándose con el mapa de calor. Esta problemática se vuelve más aparente en los casos de HSIL, debido a que este tipo de célula es la que tiene mayor relación núcleo-citoplasmática, con núcleos grandes que a penas dejan citoplasma remanente.

Por otro lado, los valores bajos de α permiten conservar la morfología celular real pero pierden la ventaja de la transparencia que nos proporciona el mapa de calor. Con un valor equilibrado como $\alpha = 0.1$, el mapa de calor se centra suavemente en el citoplasma reducido.

Es así como pequeños cambios en el valor del parámetro α producen cambios importantes en su capacidad de clasificación, con valores altos que nos ocultan información morfológica celular relevante dificultando enormemente el diagnóstico y valores demasiado bajos que pierden la ventaja que nos proporciona el mapa de calor.

El empleo de los mapas de calor (Grad-CAM) se ha empleado previamente en algunas áreas del diagnóstico por imagen. Concretamente, en el diagnóstico de citologías cervicovaginales se está empleando como alternativa a los métodos de segmentación tradicionales basados en anotaciones de imágenes médicas que son laboriosas y costosas. Los mapas de calor que separan las zonas relevantes de la imagen, se presenta como un método de explicabilidad prometedor en el campo de

diagnóstico por imagen. Se ha publicado el empleo de distintas redes enriquecidas con mapas de calor, con cifras de precisión del 94% (Sritharan et al., 2025), frente al 97,70% de precisión obtenida en nuestro estudio para $\alpha = 0.1$ en la clasificación binaria (**Tabla 5.14**). Hay otras publicaciones con resultados de precisión del 99,75% en la clasificación binaria empleando el modelo DenseNet121 enriquecido con Grad-CAM (Y. Zhang et al., 2025). Sin embargo, este estudio está realizado sobre imágenes de células individuales, eliminando grupos celulares, solapamientos y artefactos, tan frecuentes en la citología de la práctica diaria. Técnicas como la eliminación de los grupos celulares, mejora del contraste, eliminación de artefactos y mejora de la resolución (empleados por Zhang y cols.) hacen muy compleja su implantación en un escenario diagnóstico real. En nuestro proyecto hemos optado por imágenes no preprocesadas para entrenar a la red con células idénticas al escenario diagnóstico real.

6.6. MEJORA DE LA CALIDAD DE LA IMAGEN DE DISPOSITIVOS DE BAJO RENDIMIENTO Y DISEÑO DE MODELOS DE BAJO COSTE EN PAÍSES EN VÍAS DE DESARROLLO

La aplicación de la IA en entornos con recursos sanitarios limitados se plantea como una estrategia prometedora en la detección precoz del cáncer de cérvix. La escasez de patólogos experimentados y la falta de digitalización con el uso de dispositivos de bajo coste con mala calidad de la imagen, complican la implantación de los modelos de diagnóstico automatizados.

Entre nuestras estrategias de mejora encontramos el empleo de algoritmos Real-ESRGAN que mejoran la resolución y calidad de las imágenes de las células obtenidas por dispositivos de bajo coste como cámaras conectadas a microscopios o teléfonos móviles. Los resultados preliminares muestran que los modelos que entrenan a la red con imágenes con calidad mejorada, sobre todo las obtenidas de una base de datos mixta (que combina imágenes de baja, media y alta calidad)

ofrecen un mejor rendimiento que los modelos con citologías sin mejoras, alcanzando una precisión de hasta el 92,2%. El empleo de células de distintas calidades se aproxima a las citologías cervicovaginales reales, lo que podría justificar la obtención de mejores resultados en este tipo de modelos.

Del mismo modo, hemos estudiado la viabilidad de modelos TinyML como herramienta ejecutable, barata y rápida para su aplicación sobre dispositivos móviles de bajo coste. Como ventaja principal destaca que nos permite no depender de la conexión a la nube, reduciendo costes y facilitando su uso en regiones con mala conectividad como zonas rurales. El empleo de *TensorFlow lite cuantizado* ha demostrado grandes ventajas, como la reducción marcada del consumo de memoria RAM y la mejora del rendimiento móvil sin comprometer la precisión del modelo, por lo que se presenta como una solución ideal en escenarios de recursos limitados.

Además; como valor sobreañadido, el enriquecimiento de la imagen con mapas de calor en la salida de la red le añade un componente de explicabilidad que permite visualizar que regiones de la imagen han motivado la decisión diagnóstica para así mejorar la confianza en la red.

Por lo tanto; los métodos de mejora de la calidad de la imagen, el empleo de modelos bajo coste aplicables a dispositivos móviles y el empleo de métodos de explicabilidad, pueden complementar parcialmente el rol de expertos en escenarios donde el acceso a profesionales cualificados es limitado. Esta herramienta podría tener un impacto muy positivo en estos países de bajos recursos, permitiendo diagnósticos tempranos y reduciendo la morbilidad y mortalidad asociada al cáncer de cérvix.

VII – CONCLUSIONES

VII - CONCLUSIONES

- El algoritmo DL desarrollado, entrenado con imágenes celulares sin preprocesar, ha mostrado un gran rendimiento en la clasificación de la citología cervicovaginal. El uso de imágenes no preprocesadas; aunque dificulta la tarea diagnóstica al algoritmo, nos permite trabajar en condiciones más fieles a la realidad en la práctica clínica, eliminando la dependencia del procesamiento y mejorando su aplicación en entornos diagnósticos reales.

- El modelo alcanza una gran precisión diagnóstica, presentándose como una herramienta competitiva y robusta, especialmente en categorías con buena muestra de datos. La menor disponibilidad de imágenes en algunas categorías ha influido directamente en el rendimiento, lo que demuestra la importancia de equilibrar la base de datos para conservar la fiabilidad del modelo.

- El modelo híbrido DCGAN ha demostrado ser capaz de generar células artificiales de alta calidad, algunas de ellas indistinguibles de las reales por un patólogo cualificado, lo que se platea como una herramienta válida para enriquecer y aumentar las bases de datos. La creación de células artificiales nos permite compensar la baja representación de algunas categorías diagnósticas minoritarias, sin las limitaciones logísticas y éticas de recolectar muestras reales, lo que contribuye obtener modelos más compensados y robustos.

- La integración de la segmentación automática como método de interpretabilidad en el algoritmo (U-Net) ha mejorado la interpretabilidad y precisión del modelo, mejorando su rendimiento sobre todo de la clasificación binaria. Aunque el modelo binario parece más simple, ha mostrado mayor robustez y facilidad en su aplicación en la clínica diaria. Por otro lado, el modelo multiclase es más detallado y de mayor utilidad en casos clínicos complejos, pero precisa de mayor entrenamiento para alcanzar la misma fiabilidad. Factores como la desigualdad en el número de células por cada categoría diagnóstica o los artefactos de la imagen podrían afectar negativamente al rendimiento del modelo de 4 clases, que pueden ser subsanados mejorando y equilibrando la base de datos de cada categoría.

- El empleo de Grad-CAM con la elaboración de mapas de calor como método de explicabilidad, permite identificar las áreas relevantes de la imagen mostrando una alta concordancia con los criterios diagnósticos empleados por expertos. Estos mapas mejoran el rendimiento del modelo y la confianza clínica en los modelos de diagnóstico por imagen, dejando de actuar como cajas negras. El ajuste óptimo de $\alpha = 0.1$ ha demostrado un equilibrio óptimo entre la información del mapa de calor y la imagen original.

- La adaptación del modelo de diagnóstico celular a entornos de bajos recursos, mediante el uso de sistemas de mejora de imagen como Real-ESRGAN o el empleo de los modelos TinyML como modelo de bajo coste, ha demostrado ser una solución económica y viable para el diagnóstico precoz del cáncer cervical en países subdesarrollados. Estos modelos enriquecidos con métodos de explicabilidad, se plantean como una herramienta de fácil aplicación y robusta para reducir la morbimortalidad asociada al cáncer de cérvix en estos países.

VIII - LIMITACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

VIII - LIMITACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

8.1. LIMITACIONES

La limitación fundamental que encontramos al inicio del estudio, es la escasez de muestras de algunas de las categorías diagnósticas del Sistema Bethesda. En nuestro estudio preliminar y en algunas de las fases de nuestro estudio, se trató de solventar con modificaciones de la imagen con rotaciones, estiramientos o volteos, para más adelante desarrollar un sistema de creación de células artificiales con redes antagónicas. Sin embargo, las redes antagónicas presentan ciertas limitaciones en la creación de las categorías intermedias con características citológicas superponibles como ASC-US y LSIL, que se plantea como una limitación y una vía futura de mejora.

Estas categorías intermedias contienen en ocasiones una variabilidad interobservador importante, presentándose como un desafío persistente en el diagnóstico de la citología cervicovaginal, lo que remarca la necesidad de dar un enfoque objetivo y más cuantitativo a la clasificación citológica. Ponen de manifiesto la necesidad de encontrar otras estrategias diagnósticas que mejoren la consistencia diagnóstica, entre las que encontramos el uso de IA como herramienta de ayuda al diagnóstico.

El sistema Bethesda contiene otras categorías diagnósticas que no han sido incluidas en este estudio ya que; aunque pueden estar presentes, su baja frecuencia hacen poco práctica y muy farragosa su inclusión en el estudio: ASC-H (del inglés, células escamosas atípicas no se puede excluir una lesión intraepitelial escamosa de alto grado) y las atipias de células glandulares.

La categoría ASC-H indica que hay anomalías celulares mayores que un ASC-US, similares a HSIL pero que resultan insuficientes para catalogarla como tal. Son pacientes de alto riesgo que; al igual que las HSIL, requieren completar estudio

con colposcopia. La frecuencia de este tipo de lesiones es mínima, siendo con diferencia la más infrecuente de todas las categorías, lo que dificulta obtener suficiente cantidad de células necesaria para poder entrenar al modelo y obtener un modelo con robustez estadística (**Tabla 8.1**). Además, se trata de una categoría intermedia, con gran ambigüedad, difícil interpretabilidad y alta variabilidad interobservador por lo que puede influir negativamente en el entrenamiento de la red, generando ruido en el etiquetado celular, con diagnósticos poco precisos. Del mismo modo, la categoría ASC-H se solapa tanto morfológicamente (en las características celulares) como clínicamente con HSIL; requiriendo en ambos casos completar estudio con colposcopia, por considerarse ambas pacientes de alto riesgo de desarrollar cáncer.

Tabla 8.1. Estimación de la frecuencia de cada diagnóstico en la citología cervicovaginal. Son medidas estimadas obtenidas de estudios epidemiológicos y guías de cribado internacionales. Elaboración propia.

Categoría	Frecuencia estimada	Frecuencia relativa
Benigna	90-95%	1 de cada 1-1,1 citologías
ASC-US	2-5%	1 de cada 20-50 citologías
ASC-H	0,1-0,5%	1 de cada 200-1000 citologías
LSIL	1-3%	1 de cada 33-100 citologías
HSIL	0,5-1%	1 de cada 100-200 citologías

Del mismo modo, el sistema Bethesda incluye otra categoría llamada atipia de células glandulares (ACG) que hace referencia a la atipia de células que proceden del endocérvix o endometrio. Este tipo de lesiones no son el objetivo principal del cribado del cáncer de cérvix y son hallazgos secundarios que pueden observarse de forma incidental. Su carcinogénesis es independiente a la del cáncer de cérvix, el cual se origina por la infección por el HPV de células escamosas de la unión escamocolumnar. Además, al igual que ASC-H, presenta una frecuencia mínima estimada de 0,1-0,5% de todas las citologías.

Por lo tanto, y a pesar de que representa una limitación, la simplificación del modelo reduciendo el número de categorías mejora la precisión de la red y el rendimiento del modelo permitiendo una aplicación e implementación más sencilla y eficaz en la práctica diaria.

8.2. FUTURAS LÍNEAS DE INVESTIGACIÓN

Nuestro trabajo muestra resultados prometedores en todas las líneas de investigación iniciadas en las que mediante el empleo de redes antagónicas, el uso de técnicas de interpretabilidad mediante segmentación con máscaras celulares o el empleo de métodos de XIA hemos logrado mejorar los resultados para obtener un modelo robusto.

En general, aunque los resultados son excelentes, algunos de ellos son mejorables en categorías intermedias como LSIL o ASC-US, siendo las células benignas y HSIL las generadas y etiquetadas con mayor éxito por la red.

Buscando líneas futuras de investigación y revisando los criterios del sistema Bethesda, éste hace referencia al término ASC (que incluye al ASC-US) como "*cambios citológicos que sugieren SIL, pero que son cualitativamente y cuantitativamente insuficientes para establecer una interpretación definitiva*", "*cabe destacar que la categoría ASC se creó para calificar una muestra en su totalidad y no células por separado*". De esta forma se abren dos posibles líneas futuras de investigación que se tratan a continuación.

8.3. CRITERIOS DE CUANTIFICACIÓN A LA RED PARA TRATAR DE DISCERNIR ENTRE CATEGORÍAS INTERMEDIAS

Una vía de mejora de los resultados de etiquetado de LSIL y ASC-US, radica en añadir criterios de cuantificación a la red a la hora de la toma de decisiones. Actualmente la red se enfrenta a cada imagen de la célula de manera individual, clasificándola en una categoría diagnóstica. Sin embargo, en la práctica diaria el patólogo experto evalúa la muestra en conjunto con todas sus células en su totalidad y no por separado, tal y como cita el Sistema Bethesda. La distinción entre categorías intermedias como LSIL o ASC-US en ocasiones recae sobre criterios de cuantificación en los que; por ejemplo, si nos enfrentamos a muy escasas células sospechosas dentro de la muestra de la misma paciente, podemos decantarnos por ASC-US antes que por el LSIL. Es de gran importancia considerar la citología de forma global para no abusar de los diagnósticos intermedios ni sobre-diagnosticar a las pacientes con diagnósticos de LSIL si las células son cuantitativamente insuficientes.

Siguiendo esta línea de pensamiento, si la red examinara a la paciente de manera global; considerando cada una de las células que componen su citología, y fuera capaz de hacer un análisis final de todas las etiquetas obtenidas, se podrían mejorar los resultados. Revisando los resultados de nuestro estudio preliminar de Martínez-Mas y cols. en el que se entrenó a la red con imágenes reales de células, el 50% de las células LSIL se diagnosticaron erróneamente como ASC-US por la red y el 12% de las células ASC-US fueron diagnosticadas como LSIL (Martínez-Más et al., 2020) (**Tabla 8.2**). En el caso de nuestro proyecto de generar y clasificar células artificiales dentro del grupo LSIL la red clasificó erróneamente como ASC-US el 30% y dentro del grupo ASC-US la red clasificó erróneamente como LSIL el 21% (**Tabla 5.2**).

Nuestros resultados demuestran que nuestro modelo DL es una herramienta brillante en el diagnóstico de citologías cervicovaginales porque; aunque hay categorías intermedias cuyos resultados son mejorables, al considerar la citología en su conjunto aplicado a un entorno diagnóstico real, y teniendo en cuenta que las

células patológicas o displásicas no suelen ser únicas en la muestra sino que se repiten como células aisladas o en grupos, nuestra capacidad de detectar a las pacientes con displasia cervical es prácticamente del 100%. Es decir; considerando que en citopatología aunque cada muestra contiene cientos de células, el diagnóstico final de la paciente queda marcado por la etiqueta del diagnóstico más relevante o agresivo (obviando los diagnósticos de menor agresividad biológica), bastará con que la red detecte una célula patológica, será suficiente para seleccionar a aquellas pacientes con displasia cervical en riesgo de desarrollo del cáncer de cérvix. Además, considerando que hemos entrenado a la red con imágenes en un entorno real sin preprocesamiento previo, garantizamos la fácil implementación de la red a un programa de diagnóstico citológico real.

Tabla 8.2. Matriz de confusión para la clasificación multiclase de categorías individuales (H-SIL, L-SIL, ASC-US y Normal) utilizando las imágenes originales como entrada, resaltando en negrita las clasificaciones correctas en cada categoría. Tabla obtenida de la tesis doctoral del Dr. Martínez Más "Clasificación automática de tumores de ovario y citologías cervicovaginales a partir de imágenes ecográficas y microscópicas mediante su análisis con técnicas de aprendizaje automático".

Tipos celulares	**H-SIL**	**L-SIL**	**ASC-US**	**Normal**
H-SIL	**67,3 %**	8,7 %	16,0 %	8,0 %
L-SIL	22,0 %	**10,7 %**	50,0 %	14,0 %
ASC-US	27,3 %	12,0 %	**30,7 %**	30,0 %
Normal	8,7 %	8,7 %	12,7 %	**73,3 %**

8.4. CRITERIOS CUALITATIVOS QUE ANALIZA LA RED CON EL ETIQUETADO CELULAR MANUAL COMO MÉTODO DE RAZONAMIENTO DIAGNÓSTICO.

El Sistema Bethesda también hace hincapié en los criterios cualitativos celulares sobre todo en las categorías intermedias. Planteamos para una vía futura el estudio de la utilidad del etiquetado manual de células como método de enriquecimiento de la red. Para ello queremos estudiar cómo responde la red ante un entrenamiento de células etiquetadas manualmente con las características citológicas visibles en la imagen.

Tras este entrenamiento se espera que la red sea capaz de predecir automáticamente las características presentes en una imagen nueva a la que nunca se había enfrentado mediante la selección de distintas etiquetas. Si la red es capaz de relacionar los factores que justifican el diagnóstico celular, se espera mejorar la precisión y comprensión del modelo. Esto facilitaría mucho el diagnóstico ya que la red no solo diagnosticaría, sino que reflejaría por qué selecciona ese diagnóstico, describiendo de manera clara las características que están presentes en la imagen, dándole un razonamiento a su elección.

En la sección de elaboración de células artificiales con nuestro modelo híbrido DCGAN, otra vía de mejora para categorías ambiguas como ASC-US y LSIL, podría ser añadir características morfológicas a la red GAN como las descritas anteriormente, para mejorar y facilitar la elaboración de las células artificiales de estas categorías.

Por lo tanto y como punto de partida a vías futuras de investigación, se considera que tanto el enriquecimiento cuantitativo de la red como cualitativo con el etiquetado de las células, se tratan de vías de investigación prometedoras que podrían aportarle robustez al modelo.

IX - REFERENCIAS BIBLIOGRÁFICAS

IX - REFERENCIAS BIBLIOGRÁFICAS

- Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2021). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Applied Intelligence*, *51*(2), 854-864. https://doi.org/10.1007/s10489-020-01829-7
- Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digital Medicine*, *1*, 39. https://doi.org/10.1038/s41746-018-0040-6
- Acs, B., Rantalainen, M., & Hartman, J. (2020). Artificial intelligence as the next step towards precision pathology. *Journal of Internal Medicine*, *288*(1), 62-81. https://doi.org/10.1111/joim.13030
- Ahmad, J., Muhammad, K., & Baik, S. W. (2017). Medical Image Retrieval with Compact Binary Codes Generated in Frequency Domain Using Highly Reactive Convolutional Features. *Journal of Medical Systems*, *42*(2), 24. https://doi.org/10.1007/s10916-017-0875-4
- Ahmadzadeh Sarhangi, H., Beigifard, D., Farmani, E., & Bolhasani, H. (2024). Deep learning techniques for cervical cancer diagnosis based on pathology and colposcopy images. *Informatics in Medicine Unlocked*, *47*, 101503. https://doi.org/10.1016/j.imu.2024.101503
- Akca, Y., & Erkilic, S. (2023). Diagnostic utility of ThinPrep Imaging System® for detecting atypical glandular cells in cervical smear samples. *Diagnostic Cytopathology*, *51*(2), 135-139. https://doi.org/10.1002/dc.25066
- Alameda Q., F. y colaboradores. (s. f.). ASCUS y control citológico. *Revista Española de Patología*, *36*(1). Recuperado 21 de junio de 2025, de http://www.patologia.es/volumen36/vol36-num1/36-1n09.htm
- Alazawi, W. O. F., Morris, L. S., Stanley, M. A., Garrod, D. R., & Coleman, N. (2003). Altered expression of desmosomal components in high-grade squamous intraepithelial lesions of the cervix. *Virchows Archiv: An International Journal of Pathology*, *443*(1), 51-56. https://doi.org/10.1007/s00428-003-0771-9
- Albzour, N., & Lam, S. S. (2025). *Segmentation and Classification of Pap Smear Images for Cervical Cancer Detection Using Deep Learning* (arXiv:2508.17728). arXiv. https://doi.org/10.48550/arXiv.2508.17728
- Alemany L., Bayas J.M., Borruel N., Campins M., Castellsagué X., Curran A., Díaz de Heredia C., Martínez X., Moraga-Llop F.A, & Torné A. (2016). *AEPCC-Guía: VACUNACIÓN SELECTIVA FRENTE AL VIRUS DEL PAPILOMA HUMANO EN POBLACIONES DE RIESGO ELEVADO.* Publicaciones AEPCC.
- Althammer, S., Tan, T. H., Spitzmüller, A., Rognoni, L., Wiestler, T., Herz, T., Widmaier, M., Rebelatto, M. C., Kaplon, H., Damotte, D., Alifano, M., Hammond, S. A., Dieu-Nosjean, M.-C., Ranade, K., Schmidt, G., Higgs, B. W., & Steele, K. E. (2019). Automated image analysis of NSCLC biopsies to predict response to anti-PD-L1 therapy. *Journal for Immunotherapy of Cancer*, *7*, 121. https://doi.org/10.1186/s40425-019-0589-x
- Álvarez García, F. J., Iofrío De Arce, A., Álvarez Aldeán, J., Garcés-Sánchez, M., Garrote Llanos, E., Montesdeoca Melián, A., Navarro Gómez, M., Pineda Solas, V., Rivero Calle, I., Ruiz-Contreras, J., & Serrano Marchuet, P. (2024). Calendario de inmunizaciones de la Asociación Española de Pediatría: Recomendaciones 2024. *Anales de Pediatría*, *100*(1), 34-45. https://doi.org/10.1016/j.anpedi.2023.12.001
- Alves, V. A. F., Bibbo, M., Schmitt, F. C. L., Milanezi, F., & Longatto Filho, A. (2004).

Comparison of manual and automated methods of liquid-based cytology. A morphologic study. *Acta Cytologica*, *48*(2), 187-193. https://doi.org/10.1159/000326314

- Anttila, A., Kotaniemi-Talonen, L., Leinonen, M., Hakama, M., Laurila, P., Tarkkanen, J., Malila, N., & Nieminen, P. (2010). Rate of cervical cancer, severe intraepithelial neoplasia, and adenocarcinoma in situ in primary HPV DNA screening with cytology triage: Randomised study within organised screening programme. *The BMJ*, *340*, c1804. https://doi.org/10.1136/bmj.c1804
- Arbyn, M., Anttila, A., Jordan, J., Ronco, G., Schenck, U., Segnan, N., Wiener, H., Herbert, A., & von Karsa, L. (2010). European Guidelines for Quality Assurance in Cervical Cancer Screening. Second edition—Summary document. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, *21*(3), 448-458. https://doi.org/10.1093/annonc/mdp471
- Arbyn, M., Bergeron, C., Klinkhamer, P., Martin-Hirsch, P., Siebers, A. G., & Bulten, J. (2008). Liquid compared with conventional cervical cytology: A systematic review and meta-analysis. *Obstetrics and Gynecology*, *111*(1), 167-177. https://doi.org/10.1097/01.AOG.0000296488.85807.b3
- Arbyn, M., Buntinx, F., Van Ranst, M., Paraskevaidis, E., Martin-Hirsch, P., & Dillner, J. (2004). Virologic versus cytologic triage of women with equivocal Pap smears: A meta-analysis of the accuracy to detect high-grade intraepithelial neoplasia. *Journal of the National Cancer Institute*, *96*(4), 280-293. https://doi.org/10.1093/jnci/djh037
- Austin, R., & Parvathi, R. (2025). CNN based method for classifying cervical cancer cells in pap smear images. *Scientific Reports*, *15*(1), 23936. https://doi.org/10.1038/s41598-025-10009-x
- B, M.-W., R, C., & C, V. (2020). Artificial intelligence in pathology: An overview. *Diagnostic Histopathology*, *26*, 513-520.
- Bai, B., Yang, X., Li, Y., Zhang, Y., Pillar, N., & Ozcan, A. (2023). Deep learning-enabled virtual histological staining of biological samples. *Light, Science & Applications*, *12*(1), 57. https://doi.org/10.1038/s41377-023-01104-7
- Bai, J., Li, N., Ye, H., Li, X., Chen, L., Hu, J., Pang, B., Chen, X., Rao, G., Hu, Q., Liu, S., Sun, S., Li, C., Lv, X., Zeng, S., Cai, J., Cheng, S., & Liu, X. (2025). AI-assisted cervical cytology precancerous screening for high-risk population in resource-limited regions using a compact microscope. *Nature Communications*, *16*, 7429. https://doi.org/10.1038/s41467-025-62589-x
- Bengtsson, E. (1987). The measuring of cell features. *Analytical and Quantitative Cytology and Histology*, *9*(3), 212-217.
- Bengtsson, E., & Malm, P. (2014). Screening for cervical cancer using automated analysis of PAP-smears. *Computational and Mathematical Methods in Medicine*, *2014*, 842037. https://doi.org/10.1155/2014/842037
- Bergeron, C., Ordi, J., Schmidt, D., Trunk, M. J., Keller, T., Ridder, R., & European CINtec Histology Study Group. (2010). Conjunctive p16INK4a testing significantly increases accuracy in diagnosing high-grade cervical intraepithelial neoplasia. *American Journal of Clinical Pathology*, *133*(3), 395-406. https://doi.org/10.1309/AJCPXSVCDZ3D5MZM
- Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., & Lladó, X. (2019). Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: A review. *Artificial Intelligence in Medicine*, *95*, 64-81. https://doi.org/10.1016/j.artmed.2018.08.008
- Bernard, H.-U., Burk, R. D., Chen, Z., van Doorslaer, K., zur Hausen, H., & de Villiers, E.-M. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*, *401*(1), 70-79.

https://doi.org/10.1016/j.virol.2010.02.002

- Bigras, G., Wilson, J., Russell, L., Johnson, G., Morel, D., & Saddik, M. (2013). Interobserver concordance in the assessment of features used for the diagnosis of cervical atypical squamous cells and squamous intraepithelial lesions (ASC-US, ASC-H, LSIL and HSIL). *Cytopathology: Official Journal of the British Society for Clinical Cytology*, *24*(1), 44-51. https://doi.org/10.1111/j.1365-2303.2011.00930.x
- Boost, T. (2009). A comparison of screening times between the ThinPrep Imager and conventional cytology. *Diagnostic Cytopathology*, *37*(9), 661-664. https://doi.org/10.1002/dc.21069
- Bueno, G., Fernández-Carrobles, M. M., Deniz, O., & García-Rojo, M. (2016). New Trends of Emerging Technologies in Digital Pathology. *Pathobiology: Journal of Immunopathology, Molecular and Cellular Biology*, *83*(2-3), 61-69. https://doi.org/10.1159/000443482
- Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., Brogi, E., Reuter, V. E., Klimstra, D. S., & Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, *25*(8), 1301-1309. https://doi.org/10.1038/s41591-019-0508-1
- *Cáncer de cervix—SEOM: Sociedad Española de Oncología Médica © 2019*. (s. f.). Recuperado 11 de octubre de 2022, de https://seom.org/info-sobre-el-cancer/cervix?start=1
- *Cancer today*. (s. f.). Recuperado 15 de enero de 2023, de http://gco.iarc.fr/today/home
- Cantley, R. L., Jing, X., Smola, B., Hao, W., Harrington, S., & Pantanowitz, L. (2024). Validation of AI-assisted ThinPrep® Pap test screening using the GeniusTM Digital Diagnostics System. *Journal of Pathology Informatics*, *15*, 100391. https://doi.org/10.1016/j.jpi.2024.100391
- Castellsagué, X. (2008). Natural history and epidemiology of HPV infection and cervical cancer. *Gynecologic Oncology*, *110*(3 Suppl 2), S4-7. https://doi.org/10.1016/j.ygyno.2008.07.045
- Castellsagué, X., Bosch, F. X., & Muñoz, N. (2002). Environmental co-factors in HPV carcinogenesis. *Virus Research*, *89*(2), 191-199. https://doi.org/10.1016/s0168-1702(02)00188-0
- Castellsagué, X., Iftner, T., Roura, E., Vidart, J. A., Kjaer, S. K., Bosch, F. X., Muñoz, N., Palacios, S., San Martin Rodriguez, M., Serradell, L., Torcel-Pagnon, L., Cortes, J., & CLEOPATRE Spain Study Group. (2012). Prevalence and genotype distribution of human papillomavirus infection of the cervix in Spain: The CLEOPATRE study. *Journal of Medical Virology*, *84*(6), 947-956. https://doi.org/10.1002/jmv.23282
- Castellsagué, X., Naud, P., Chow, S.-N., Wheeler, C. M., Germar, M. J. V., Lehtinen, M., Paavonen, J., Jaisamrarn, U., Garland, S. M., Salmerón, J., Apter, D., Kitchener, H., Teixeira, J. C., Skinner, S. R., Limson, G., Szarewski, A., Romanowski, B., Aoki, F. Y., Schwarz, T. F., … Baril, L. (2014). Risk of newly detected infections and cervical abnormalities in women seropositive for naturally acquired human papillomavirus type 16/18 antibodies: Analysis of the control arm of PATRICIA. *The Journal of Infectious Diseases*, *210*(4), 517-534. https://doi.org/10.1093/infdis/jiu139
- Castillo, M., Astudillo, A., Clavero, O., Velasco, J., Ibáñez, R., & de Sanjosé, S. (2016). Poor Cervical Cancer Screening Attendance and False Negatives. A Call for Organized Screening. *PLoS ONE*, *11*(8), e0161403. https://doi.org/10.1371/journal.pone.0161403
- Cervical cancer. (2022, febrero 22). *World Health Organization*. https://www.who.int/es/news-room/fact-sheets/detail/cervical-cancer
- Chang, H. Y., Jung, C. K., Woo, J. I., Lee, S., Cho, J., Kim, S. W., & Kwak, T.-Y. (2019). Artificial Intelligence in Pathology. *Journal of Pathology and Translational Medicine*, *53*(1), 1-12. https://doi.org/10.4132/jptm.2018.12.16

- Chantziantoniou, N. (2023). BestCyte® primary screening of 500 ThinPrep Pap Test thin-layers: 3 Cytologists' Interobserver diagnostic concordance with predicate manual microscopy relative to Truth Reference diagnoses defining NILM, ASCUS+, LSIL+, and ASCH+ thresholds for specificity, sensitivity, and equivalency grading. *Journal of Pathology Informatics*, *14*, 100182. https://doi.org/10.1016/j.jpi.2022.100182
- Chen, X., Yu, J., Cheng, S., Geng, X., Liu, S., Han, W., Hu, J., Chen, L., Liu, X., & Zeng, S. (2021). An unsupervised style normalization method for cytopathology images. *Computational and Structural Biotechnology Journal*, *19*. https://doi.org/10.1016/j.csbj.2021.06.025
- Chivukula, M., Saad, R. S., Elishaev, E., White, S., Mauser, N., & Dabbs, D. J. (2007). Introduction of the Thin Prep Imaging System™ (TIS): Experience in a high volume academic practice. *CytoJournal*, *4*, 6. https://doi.org/10.1186/1742-6413-4-6
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, *35*(1), 53-65. IEEE Signal Processing Magazine. https://doi.org/10.1109/MSP.2017.2765202
- Croitoru, F.-A., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion Models in Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(9), 10850-10869. https://doi.org/10.1109/TPAMI.2023.3261988
- Crowell, E. F., Bazin, C., Saunier, F., Brixtel, R., Caillot, Y., Lesner, B., Toutain, M., Ferreri, C., Garcia, I., Mathieu, M.-C., Vaussanvin, J., Depardon, J., & Renouf, A. (2019). CytoProcessorTM: A New Cervical Cancer Screening System for Remote Diagnosis. *Acta Cytologica*, *63*(3), 215-223. https://doi.org/10.1159/000497111
- Cuestas E. (2008). *El Premio Nobel de fisiolog ía o medicina*.
- Darragh, T. M., Colgan, T. J., Thomas Cox, J., Heller, D. S., Henry, M. R., Luff, R. D., McCalmont, T., Nayar, R., Palefsky, J. M., Stoler, M. H., Wilkinson, E. J., Zaino, R. J., Wilbur, D. C., & Members of the LAST Project Work Groups. (2013). The Lower Anogenital Squamous Terminology Standardization project for HPV-associated lesions: Background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *International Journal of Gynecological Pathology: Official Journal of the International Society of Gynecological Pathologists*, *32*(1), 76-115. https://doi.org/10.1097/PGP.0b013e31826916c7
- de Bekker-Grob, E. W., de Kok, I. M. C. M., Bulten, J., van Rosmalen, J., Vedder, J. E. M., Arbyn, M., Klinkhamer, P. J. J. M., Siebers, A. G., & van Ballegooijen, M. (2012). Liquid-based cervical cytology using ThinPrep technology: Weighing the pros and cons in a cost-effectiveness analysis. *Cancer Causes & Control: CCC*, *23*(8), 1323-1331. https://doi.org/10.1007/s10552-012-0011-1
- De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., … Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, *24*(9), 1342-1350. https://doi.org/10.1038/s41591-018-0107-6
- Diniz, D., Rezende, M., Campos, A., Carneiro, C., Luz, E., Moreira, G., Ushizima, D., Medeiros, F., & Souza, M. (2021). A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification. *Journal of Imaging*, *7*, 111. https://doi.org/10.3390/jimaging7070111
- Dong, N., Zhao, L., Wu, C.-H., & Chang, J. F. (2020). Inception v3 based cervical cell classification combined with artificially extracted features. *Applied Soft Computing*, *93*,

106311. https://doi.org/10.1016/j.asoc.2020.106311

- Ehteshami Bejnordi, B., Mullooly, M., Pfeiffer, R. M., Fan, S., Vacek, P. M., Weaver, D. L., Herschorn, S., Brinton, L. A., van Ginneken, B., Karssemeijer, N., Beck, A. H., Gierach, G. L., van der Laak, J. A. W. M., & Sherman, M. E. (2018). Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Modern Pathology*, *31*(10), 1502-1512. https://doi.org/10.1038/s41379-018-0073-z
- Elishaev, E., Harinath, L., Ye, Y., Matsko, J., Colaizzi, A., Wharton, S., Bhargava, R., Pantanowitz, L., Hanna, M. G., Harrington, S., & Zhao, C. (2025). Assessment of the efficacy and accuracy of cervical cytology screening with the Hologic Genius Digital Diagnostics System. *Cancer Cytopathology*, *133*(7), e70022. https://doi.org/10.1002/cncy.70022
- Ertosun, M. G., & Rubin, D. L. (2015). Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2015*, 1899-1908.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118. https://doi.org/10.1038/nature21056
- Fuchs, T. J., & Buhmann, J. M. (2011). Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, *35*(7-8), 515-530. https://doi.org/10.1016/j.compmedimag.2011.02.006
- Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151. https://doi.org/10.1016/j.engappai.2022.105151
- García-Laencina, P. J., Abreu, P. H., Abreu, M. H., & Afonoso, N. (2015). Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. *Computers in Biology and Medicine*, *59*, 125-133. https://doi.org/10.1016/j.compbiomed.2015.02.006
- García-Laencina, P. J., Rodríguez-Bermúdez, G., & Roca-Dorda, J. (2014). *Feed-Forward Neural Network Architectures Based on Extreme Learning Machine for Parkinson's Disease Diagnosis* (L. M. Roa Romero, Ed.; Vol. 41, pp. 746-749). Springer International Publishing. https://doi.org/10.1007/978-3-319-00846-2_185
- Garud, H., Karri, S. P. K., Sheet, D., Chatterjee, J., Mahadevappa, M., Ray, A. K., Ghosh, A., & Maity, A. K. (2017). *High-Magnification Multi-Views Based Classification of Breast Fine Needle Aspiration Cytology Cell Samples Using Fusion of Decisions From Deep Convolutional Networks*. 76-81. https://openaccess.thecvf.com/content_cvpr_2017_workshops/w8/html/Garud_High-Magnification_Multi-Views_Based_CVPR_2017_paper.html
- Giuliano, A. R., Anic, G., & Nyitray, A. G. (2010). Epidemiology and pathology of HPV disease in males. *Gynecologic Oncology*, *117*(2 Suppl), S15-19. https://doi.org/10.1016/j.ygyno.2010.01.026
- Godinez, W. J., Hossain, I., Lazic, S. E., Davies, J. W., & Zhang, X. (2017). A multi-scale convolutional neural network for phenotyping high-content cellular images. *Bioinformatics (Oxford, England)*, *33*(13), 2010-2019. https://doi.org/10.1093/bioinformatics/btx069
- Goodfellow I, Bengio Y, Courville A, & Bengio Y. (2016). *Deep learning, tomo 1.* MIT press Cambridge.
- Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., & Liu, J. (2019). CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE*

Transactions on Medical Imaging, *38*(10), 2281-2292. IEEE Transactions on Medical Imaging. https://doi.org/10.1109/TMI.2019.2903562

- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, *316*(22), 2402-2410. https://doi.org/10.1001/jama.2016.17216
- Guo, Z., Shen, Y., Wan, S., Shang, W.-L., & Yu, K. (2022). Hybrid Intelligence-Driven Medical Image Recognition for Remote Patient Diagnosis in Internet of Medical Things. *IEEE Journal of Biomedical and Health Informatics*, *26*(12), 5817-5828. https://doi.org/10.1109/JBHI.2021.3139541
- Gupta, R., Kumar, N., Bansal, S., Singh, S., Sood, N., & Gupta, S. (2023). Artificial Intelligence-driven Digital Cytology-based Cervical Cancer Screening: Is the Time Ripe to Adopt This Disruptive Technology in Resource-constrained Settings? A Literature Review. *Journal of Digital Imaging*, *36*(4), 1643-1652. https://doi.org/10.1007/s10278-023-00821-0
- Hakama, M., & Räsänen-Virtanen, U. (1976). Effect of a mass screening program on the risk of cervical cancer. *American Journal of Epidemiology*, *103*(5), 512-517. https://doi.org/10.1093/oxfordjournals.aje.a112253
- Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism: Clinical and Experimental*, *69S*, S36-S40. https://doi.org/10.1016/j.metabol.2017.01.011
- Hartwig, S., Baldauf, J.-J., Dominiak-Felden, G., Simondon, F., Alemany, L., de Sanjosé, S., & Castellsagué, X. (2015). Estimation of the epidemiological burden of HPV-related anogenital cancers, precancerous lesions, and genital warts in women and men in Europe: Potential additional benefit of a nine-valent second generation HPV vaccine compared to first generation HPV vaccines. *Papillomavirus Research*, *1*, 90-100. https://doi.org/10.1016/j.pvr.2015.06.003
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. https://doi.org/10.1109/CVPR.2016.90
- He, L., Wang, S., Chen, C., Wang, Y., Fan, Q., Chu, C., Fan, L., & Xu, J. (2025). Network Occlusion Sensitivity Analysis Identifies Regional Contributions to Brain Age Prediction. *Human Brain Mapping*, *46*(8), e70239. https://doi.org/10.1002/hbm.70239
- Heckerman, D. E., Horvitz, E. J., & Nathwani, B. N. (2018). Toward Normative Expert Systems: Part I The Pathfinder Project. *Methods of Information in Medicine*, *31*, 90-105. https://doi.org/10.1055/s-0038-1634867
- Heckerman, D. E., & Nathwani, B. N. (1992). Toward normative expert systems: Part II. Probability-based representations for efficient knowledge acquisition and inference. *Methods of Information in Medicine*, *31*(2), 106-116.
- Hekler, A., Utikal, J. S., Enk, A. H., Solass, W., Schmitt, M., Klode, J., Schadendorf, D., Sondermann, W., Franklin, C., Bestvater, F., Flaig, M. J., Krahl, D., von Kalle, C., Fröhling, S., & Brinker, T. J. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *European Journal of Cancer (Oxford, England: 1990)*, *118*, 91-96. https://doi.org/10.1016/j.ejca.2019.06.012
- Holowaty, P., Miller, A. B., Rohan, T., & To, T. (1999). Natural history of dysplasia of the uterine cervix. *Journal of the National Cancer Institute*, *91*(3), 252-258. https://doi.org/10.1093/jnci/91.3.252
- Hutchinson, M. L., Zahniser, D. J., Sherman, M. E., Herrero, R., Alfaro, M., Bratti, M. C., Hildesheim, A., Lorincz, A. T., Greenberg, M. D., Morales, J., & Schiffman, M. (1999). Utility of liquid-based cytology for cervical carcinoma screening: Results of a

population-based study conducted in a region of Costa Rica with a high incidence of cervical carcinoma. *Cancer*, *87*(2), 48-55. https://doi.org/10.1002/(sici)1097-0142(19990425)87:2<48::aid-cncr2>3.0.co;2-d

- Ibáñez, R., Alejo, M., Combalia, N., Tarroch, X., Autonell, J., Codina, L., Culubret, M., Bosch, F. X., & de Sanjosé, S. (2015). Underscreened Women Remain Overrepresented in the Pool of Cervical Cancer Cases in Spain: A Need to Rethink the Screening Interventions. *BioMed Research International*, *2015*, 605375. https://doi.org/10.1155/2015/605375
- Ismail, S. M., Colclough, A. B., Dinnen, J. S., Eakins, D., Evans, D. M., Gradwell, E., O'Sullivan, J. P., Summerell, J. M., & Newcombe, R. G. (1989). Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia. *BMJ : British Medical Journal*, *298*(6675), 707-710.
- Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, *7*, 29. https://doi.org/10.4103/2153-3539.186902
- Jiang, P., Li, X., Shen, H., Chen, Y., Wang, L., Chen, H., Feng, J., & Liu, J. (2023). A systematic review of deep learning-based cervical cytology screening: From cell identification to whole slide image analysis. *Artificial Intelligence Review*, *56*(2), 2687-2758. https://doi.org/10.1007/s10462-023-10588-z
- Jung, Y., Kim, T., Han, M.-R., Kim, S., Kim, G., Lee, S., & Choi, Y. J. (2022). Ovarian tumor diagnosis using deep convolutional neural networks and a denoising convolutional autoencoder. *Scientific Reports*, *12*(1), 17024. https://doi.org/10.1038/s41598-022-20653-2
- Jurtz, V. I., Johansen, A. R., Nielsen, M., Almagro Armenteros, J. J., Nielsen, H., Sønderby, C. K., Winther, O., & Sønderby, S. K. (2017). An introduction to deep learning on biological sequence data: Examples and solutions. *Bioinformatics (Oxford, England)*, *33*(22), 3685-3690. https://doi.org/10.1093/bioinformatics/btx531
- Ker, J., Bai, Y., Lee, H. Y., Rao, J., & Wang, L. (2019). Automated brain histology classification using machine learning. *Journal of Clinical Neuroscience: Official Journal of the Neurosurgical Society of Australasia*, *66*, 239-245. https://doi.org/10.1016/j.jocn.2019.05.019
- Khozaimi, A., Darti, I., Anam, S., & Kusumawinahyu, W. M. (2025). Advanced cervical cancer classification: Enhancing pap smear images with hybrid PMD Filter-CLAHE. *Indonesian Journal of Electrical Engineering and Computer Science*, *39*(1), 644. https://doi.org/10.11591/ijeecs.v39.i1.pp644-655
- Kitchener, H. C., Blanks, R., Cubie, H., Desai, M., Dunn, G., Legood, R., Gray, A., Sadique, Z., Moss, S., & MAVARIC Trial Study Group. (2011). MAVARIC - a comparison of automation-assisted and manual cervical screening: A randomised controlled trial. *Health Technology Assessment (Winchester, England)*, *15*(3), iii-iv, ix-xi, 1-170. https://doi.org/10.3310/hta15030
- Koliopoulos, G., Nyaga, V. N., Santesso, N., Bryant, A., Martin-Hirsch, P. P., Mustafa, R. A., Schünemann, H., Paraskevaidis, E., & Arbyn, M. (2017). Cytology versus HPV testing for cervical cancer screening in the general population. *The Cochrane Database of Systematic Reviews*, *8*(8), CD008587. https://doi.org/10.1002/14651858.CD008587.pub2
- Kolstad, P. (1970). Diagnosis and Management of Precancerous Lesions of the Cervix Uteri. *International Journal of Gynecology & Obstetrics*, *8*(4P2), 551-560. https://doi.org/10.1002/j.1879-3479.1970.tb00171.x
- Koss, L. G., Lin, E., Schreiber, K., Elgert, P., & Mango, L. (1994). Evaluation of the PAPNET cytologic screening system for quality control of cervical smears. *American Journal of Clinical Pathology*, *101*(2), 220-229. https://doi.org/10.1093/ajcp/101.2.220
- Koss, L. G., Sherman, M. E., Cohen, M. B., Anes, A. R., Darragh, T. M., Lemos, L. B.,

McClellan, B. J., Rosenthal, D. L., Keyhani-Rofagha, S., Schreiber, K., & Valente, P. T. (1997). Significant reduction in the rate of false-negative cervical smears with neural network-based technology (PAPNET Testing System). *Human Pathology*, *28*(10), 1196-1203. https://doi.org/10.1016/s0046-8177(97)90258-6

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. https://doi.org/10.1038/nature14539
- LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010). Convolutional networks and applications in vision. *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, 253-256. https://doi.org/10.1109/ISCAS.2010.5537907
- Lee, K. R., Ashfaq, R., Birdsong, G. G., Corkill, M. E., McIntosh, K. M., & Inhorn, S. L. (1997). Comparison of conventional Papanicolaou smears and a fluid-based, thin-layer system for cervical cancer screening. *Obstetrics and Gynecology*, *90*(2), 278-284. https://doi.org/10.1016/S0029-7844(97)00228-7
- Lew, M., Wilbur, D. C., & Pantanowitz, L. (2021). Computational Cytology: Lessons Learned from Pap Test Computer-Assisted Screening. *Acta Cytologica*, *65*(4), 286-300. https://doi.org/10.1159/000508629
- *Libro Blanco Anat. Patologica-2013*. (s. f.).
- Libro Blanco de la Anatomía Patológica en España 2021. (2021). *de la*, 331.
- Luisa Sánchez Lorenzo. (2023, abril 14). Cáncer de cérvix. *SEOM*. https://seom.org/174-Informaci%C3%B3n%20al%20P%C3%BAblico%20-%20Patolog%C3%ADas/cancer-de-cervix#:~:text=En%20Espa%C3%B1a%20el%20c%C3%A1ncer%20de,casos%2F100.000%20mujeres%20%2F%20a%C3%B1o.
- Ma, J., Liu, S., Cheng, S., Chen, R., Liu, X., Chen, L., & Zeng, S. (2022). STSRNet: Self-Texture Transfer Super-Resolution and Refocusing Network. *IEEE Transactions on Medical Imaging*, *41*(2), 383-393. https://doi.org/10.1109/TMI.2021.3112923
- Marcial García Rojo. (2021). *Aplicaciones de deep learning, un método de inteligencia artificial, en anatomía patológica* (Libro de la Blanco Anatomía Patológica en España 2021).
- Martínez, J. C. (2006). Nomenclatura y diagnóstico de las lesiones intraepiteliales cervicales. *Medimay*, *12*(1), Article 1.
- Martínez-Más, J., Bueno-Crespo, A., Martínez-España, R., Remezal-Solano, M., Ortiz-González, A., Ortiz-Reina, S., & Martínez-Cendán, J.-P. (2020). Classifying Papanicolaou cervical smears through a cell merger approach by deep learning technique. *Expert Systems with Applications*, *160*, 113707. https://doi.org/10.1016/j.eswa.2020.113707
- McCredie, M. R. E., Sharples, K. J., Paul, C., Baranyai, J., Medley, G., Jones, R. W., & Skegg, D. C. G. (2008). Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: A retrospective cohort study. *The Lancet. Oncology*, *9*(5), 425-434. https://doi.org/10.1016/S1470-2045(08)70103-7
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115-133. https://doi.org/10.1007/BF02478259
- Melnikow, J., Nuovo, J., Willan, A. R., Chan, B. K., & Howell, L. P. (1998). Natural history of cervical squamous intraepithelial lesions: A meta-analysis. *Obstetrics and Gynecology*, *92*(4 Pt 2), 727-735. https://doi.org/10.1016/s0029-7844(98)00245-2
- Mensah, S. O., Neubauer, J., Ayhan, M. S., Djoumessi, K., Koch, L., Uzel, M. M., Gelisken, F., & Berens, P. (2025). *Clinically Interpretable Deep Learning via Sparse BagNets for Epiretinal Membrane and Related Pathology Detection* (p. 2025.06.05.25329045). medRxiv. https://doi.org/10.1101/2025.06.05.25329045

- Miller, D. D., & Brown, E. W. (2018). Artificial Intelligence in Medical Practice: The Question to the Answer? *The American Journal of Medicine*, *131*(2), 129-133. https://doi.org/10.1016/j.amjmed.2017.10.035
- Mills, S. (2019). *Histology for Pathologists* (5.ª ed.). LWW.
- Moeskops, P., Viergever, M., Mendrik, A., Vries, L., Benders, M., & Isgum, I. (2016). Automatic Segmentation of MR Brain Images With a Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, *35*, 1252-1261. https://doi.org/10.1109/TMI.2016.2548501
- Morales Martínez, Á., Blanco Rodríguez, L., Morales Martínez, C., & Tejuca Somoano, S. (2015). Cribado de cáncer de cérvix: ¿merece la pena la búsqueda activa? *Atención Primaria*, *47*(10), 653-658. https://doi.org/10.1016/j.aprim.2014.12.006
- Moscicki, A.-B., Ma, Y., Wibbelsman, C., Darragh, T. M., Powers, A., Farhat, S., & Shiboski, S. (2010). Rate of and risks for regression of cervical intraepithelial neoplasia 2 in adolescents and young women. *Obstetrics and Gynecology*, *116*(6), 1373-1380. https://doi.org/10.1097/AOG.0b013e3181fe777f
- Mothkur, R., Soubhagyalakshmi, P., & B, S. C. (2025). Grad-CAM based Visualization for Interpretable Lung Cancer Categorization using Deep CNN Models. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, *7*(3), 567-580. https://doi.org/10.35882/jeeemi.v7i3.690
- Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., & Hajirasouliha, I. (2018). Breast Cancer Histopathological Image Classification: A Deep Learning Approach. *bioRxiv*, 242818. https://doi.org/10.1101/242818
- Nayar, R., & Wilbur, D. C. (2015). The Pap test and Bethesda 2014. *Cancer Cytopathology*, *123*(5), 271-281. https://doi.org/10.1002/cncy.21521
- Nayar, R., & Wilbur, D. C. (2017). The Bethesda System for Reporting Cervical Cytology: A Historical Perspective. *Acta Cytologica*, *61*(4-5), 359-372. https://doi.org/10.1159/000477556
- Niazi, M. K. K., Parwani, A. V., & Gurcan, M. N. (2019). Digital pathology and artificial intelligence. *The Lancet. Oncology*, *20*(5), e253-e261. https://doi.org/10.1016/S1470-2045(19)30154-8
- Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, *452*, 48-62. https://doi.org/10.1016/j.neucom.2021.03.091
- Nott, J. P., Bonney, E. A., Pickering, J. D., & Simpson, N. A. B. (2016). The structure and function of the cervix during pregnancy. *Translational Research in Anatomy*, *2*, 1-7. https://doi.org/10.1016/j.tria.2016.02.001
- Ong, S. H., Jin, X. C., Jayasooriah, & Sinniah, R. (1996). Image analysis of tissue sections. *Computers in Biology and Medicine*, *26*(3), 269-279. https://doi.org/10.1016/0010-4825(96)00004-2
- Ordi, J., Garcia, S., del Pino, M., Landolfi, S., Alonso, I., Quintó, L., & Torné, A. (2009). P16 INK4a immunostaining identifies occult CIN lesions in HPV-positive women. *International Journal of Gynecological Pathology: Official Journal of the International Society of Gynecological Pathologists*, *28*(1), 90-97. https://doi.org/10.1097/PGP.0b013e31817e9ac5
- Ortiz-González, A., Martínez-España, R., Morales-García, J., Imbernón, B., Martínez-Más, J., Álvarez, M. A., David Romero, O., Pedro Martínez-Cendán, J., & Bueno-Crespo, A. (2025). Segmentation Techniques Applied to CNNs for Cervical Cancer Classification. *IEEE Access*, *13*, 74678-74688. https://doi.org/10.1109/ACCESS.2025.3562762
- Pangarkar, M. A. (2022). The Bethesda System for reporting cervical cytology. *CytoJournal*, *19*, 28. https://doi.org/10.25259/CMAS_03_07_2021

- Papanicolaou, G. N., & Traut, H. F. (1997). The diagnostic value of vaginal smears in carcinoma of the uterus. 1941. *Archives of Pathology & Laboratory Medicine*, *121*(3), 211-224.
- Parra, V., Espinosa, A., González, A., & Biosca, J. (2020). Patología digital para técnicos especialistas en anatomía patológica (TEAP). *Archivos de Patologia*, *1*. https://doi.org/10.47579/AP.20.04.0029
- Patten, S. F., Lee, J. S., & Nelson, A. C. (1996). NeoPath, Inc. NeoPath AutoPap 300 Automatic Pap Screener System. *Acta Cytologica*, *40*(1), 45-52. https://doi.org/10.1159/000333584
- Patten, S. F., Lee, J. S., Wilbur, D. C., Bonfiglio, T. A., Colgan, T. J., Richart, R. M., Cramer, H., & Moinuddin, S. (1997). The AutoPap 300 QC System multicenter clinical trials for use in quality control rescreening of cervical smears: I. A prospective intended use study. *Cancer*, *81*(6), 337-342. https://doi.org/10.1002/(sici)1097-0142(19971225)81:6<337::aid-cncr7>3.0.co;2-i
- Prendiville, W., & Sankaranarayanan, R. (2017). Anatomy of the uterine cervix and the transformation zone. En *Colposcopy and Treatment of Cervical Precancer*. International Agency for Research on Cancer. https://www.ncbi.nlm.nih.gov/books/NBK568392/
- Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Wu, X., Li, X., & Wang, Q. (2021). DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. *Computers in Biology and Medicine*, *136*, 104649. https://doi.org/10.1016/j.compbiomed.2021.104649
- Rakha, E. A., Toss, M., Shiino, S., Gamble, P., Jaroensri, R., Mermel, C. H., & Chen, P.-H. C. (2021). Current and future applications of artificial intelligence in pathology: A clinical perspective. *Journal of Clinical Pathology*, *74*(7), 409-414. https://doi.org/10.1136/jclinpath-2020-206908
- Rasool, N., Wani, N. A., Bhat, J. I., Saharan, S., Sharma, V. K., Alsulami, B. S., Alsharif, H., & Lytras, M. D. (2025). CNN-TumorNet: Leveraging explainability in deep learning for precise brain tumor diagnosis on MRI images. *Frontiers in Oncology*, *15*, 1554559. https://doi.org/10.3389/fonc.2025.1554559
- Rivera Rolon, M. D. M., Gustafson, E., Cole, R., Matos, J., Hicken, K., Hicks, J., Cahoon, B., de Socarraz, M., & Santa-Rosario, J. C. (2025). Implementing 100% quality control in a cervical cytology workflow using whole slide images and artificial intelligence provided by the Techcyte SureView™ System. *Cancer Cytopathology*, *133*(6), e70019. https://doi.org/10.1002/cncy.70019
- Robertson, S., Azizpour, H., Smith, K., & Hartman, J. (2018). Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. *Translational Research: The Journal of Laboratory and Clinical Medicine*, *194*, 19-35. https://doi.org/10.1016/j.trsl.2017.10.010
- Rojo, M. G. (2019). Inteligencia artificial en Anatomía Patológica. *Revista Española de Patología*, *52*(4), 205-207. https://doi.org/10.1016/j.patol.2019.09.001
- Ronco, G., Cuzick, J., Pierotti, P., Cariaggi, M. P., Dalla Palma, P., Naldoni, C., Ghiringhello, B., Giorgi-Rossi, P., Minucci, D., Parisio, F., Pojer, A., Schiboni, M. L., Sintoni, C., Zorzi, M., Segnan, N., & Confortini, M. (2007). Accuracy of liquid based versus conventional cytology: Overall results of new technologies for cervical cancer screening: randomised controlled trial. *BMJ (Clinical Research Ed.)*, *335*(7609), 28. https://doi.org/10.1136/bmj.39196.740995.BE
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. En N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234-241). Springer International Publishing. https://doi.org/10.1007/978-3-

319-24574-4_28

- Ross MH, Wojciech P. (2013). *Histología*.
- Rouvière, H. (2005). *Anatomía humana descriptiva, topográfica y funcional: Vol. 2. Tronco* (11.ª ed.). Elsevier.
- Saltz, J. H. (2000). Digital pathology—The big picture. *Human Pathology*, *31*(7), 779-780.
- Saslow, D., Solomon, D., Lawson, H. W., Killackey, M., Kulasingam, S. L., Cain, J., Garcia, F. A. R., Moriarty, A. T., Waxman, A. G., Wilbur, D. C., Wentzensen, N., Downs, L. S., Spitzer, M., Moscicki, A.-B., Franco, E. L., Stoler, M. H., Schiffman, M., Castle, P. E., Myers, E. R., & ACS-ASCCP-ASCP Cervical Cancer Guideline Committee. (2012). American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA: A Cancer Journal for Clinicians*, *62*(3), 147-172. https://doi.org/10.3322/caac.21139
- Sellors, J. W., & Sankaranarayanan, R. (s. f.). (2003). *La colposcopia y el tratamiento de la neoplasia intraepitelial cervical:* 140.
- Sendagorta-Cudós, E., Burgos-Cibrián, J., & Rodríguez-Iglesias, M. (2019). Infecciones genitales por el virus del papiloma humano. *Enfermedades Infecciosas y Microbiología Clínica*, *37*(5), 324-334. https://doi.org/10.1016/j.eimc.2019.01.010
- Sha, Y., Zhang, Q., Zhai, X., Hou, M., Lu, J., Meng, W., Wang, Y., Li, K., & Ma, J. (2024). CerviFusionNet: A multi-modal, hybrid CNN-transformer-GRU model for enhanced cervical lesion multi-classification. *iScience*, *27*(12), 111313. https://doi.org/10.1016/j.isci.2024.111313
- Shao, J., Chen, S., Zhou, J., Zhu, H., Wang, Z., & Brown, M. (2023). Application of U-Net and Optimized Clustering in Medical Image Segmentation: A Review. *Computer Modeling in Engineering & Sciences*, *136*(3), 2173-2219. https://doi.org/10.32604/cmes.2023.025499
- Shortliffe, E. H. (1977). Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 66-69.
- Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv. https://doi.org/10.48550/arXiv.1409.1556
- Solomon, D., Davey, D., Kurman, R., Moriarty, A., O'Connor, D., Prey, M., Raab, S., Sherman, M., Wilbur, D., Wright, T., Young, N., Forum Group Members, & Bethesda 2001 Workshop. (2002). The 2001 Bethesda System: Terminology for reporting results of cervical cytology. *JAMA*, *287*(16), 2114-2119. https://doi.org/10.1001/jama.287.16.2114
- Sørbye, S. W., Suhrke, P., Revå, B. W., Berland, J., Maurseth, R. J., & Al-Shibli, K. (2017). Accuracy of cervical cytology: Comparison of diagnoses of 100 Pap smears read by four pathologists at three hospitals in Norway. *BMC Clinical Pathology*, *17*, 18. https://doi.org/10.1186/s12907-017-0058-8
- Sritharan, N., Gnanavel, N., Inparaj, P., Meedeniya, D., & Yogarajah, P. (2025). Explainable Artificial Intelligence Driven Segmentation for Cervical Cancer Screening. *IEEE Access*, *13*, 71306-71322. https://doi.org/10.1109/ACCESS.2025.3561178
- Stoler, M. H., Schiffman, M., & Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. (2001). Interobserver reproducibility of cervical cytologic and histologic interpretations: Realistic estimates from the ASCUS-LSIL Triage Study. *JAMA*, *285*(11), 1500-1505. https://doi.org/10.1001/jama.285.11.1500
- Stoler, M. H., Wright, T. C., Ferenczy, A., Ranger-Moore, J., Fang, Q., Kapadia, M., & Ridder, R. (2018). Routine Use of Adjunctive p16 Immunohistochemistry Improves Diagnostic Agreement of Cervical Biopsy Interpretation: Results From the CERTAIN

Study. *The American Journal of Surgical Pathology*, *42*(8), 1001-1009. https://doi.org/10.1097/PAS.0000000000001072

- Suara, S., Jha, A., Sinha, P., & Sekh, A. A. (2024). *Is Grad-CAM Explainable in Medical Images?* (Vol. 2009, pp. 124-135). https://doi.org/10.1007/978-3-031-58181-6_11
- Sun, C., Guo, S., Zhang, H., Li, J., Chen, M., Ma, S., Jin, L., Liu, X., Li, X., & Qian, X. (2017). Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs. *Artificial Intelligence in Medicine*, *83*, 58-66. https://doi.org/10.1016/j.artmed.2017.03.008
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209-249. https://doi.org/10.3322/caac.21660
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). *Rethinking the Inception Architecture for Computer Vision* (arXiv:1512.00567). arXiv. https://doi.org/10.48550/arXiv.1512.00567
- Taha, B., Dias, J., & Werghi, N. (2017). *Classification of Cervical-Cancer Using Pap-Smear Images: A Convolutional Neural Network Approach* (M. Valdés Hernández & V. González-Castro, Eds.; Vol. 723, pp. 261-272). Springer International Publishing. https://doi.org/10.1007/978-3-319-60964-5_23
- Tareef, A., Song, Y., Cai, W., Huang, H., Chang, H., Wang, Y., Fulham, M., Feng, D., & Chen, M. (2017). Automatic segmentation of overlapping cervical smear cells based on local distinctive features and guided shape deformation. *Neurocomputing*, *221*, 94-107. https://doi.org/10.1016/j.neucom.2016.09.070
- Tareef, A., Song, Y., Huang, H., Wang, Y., Feng, D., Chen, M., & Cai, W. (2017). Optimizing the cervix cytological examination based on deep learning and dynamic shape modeling. *Neurocomputing*, *248*, 28-40. https://doi.org/10.1016/j.neucom.2017.01.093
- Teresa Queiro Verdes & Jeannette Puñal Riobóo. (2013, enero). *Métodos automatizados de lectura de citología cervical uterina*. Axencia de Avaliación de Tecnoloxías Sanitarias de Galicia, Consellería de Sanidade. https://www.sergas.es/Docs/Avalia-t/avalia-t201301Lecturaautomatizada.pdf
- Thorstenson, S., Molin, J., & Lundström, C. (2014). Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006-2013. *Journal of Pathology Informatics*, *5*(1), 14. https://doi.org/10.4103/2153-3539.129452
- Tolles, W. E. (1955). The cytoanalyzer-an example of physics in medical research. *Transactions of the New York Academy of Sciences*, *17*(3), 250-256. https://doi.org/10.1111/j.2164-0947.1955.tb01204.x
- Torné A., Andía D., Bruni L., Centeno C., Coronado P, Cruz Quílez J, de la Fuente J, de Sanjosé S, Granados R, Ibáñez R, Lloveras B, Lubrano A Matías Guiu X, Medina N, Ordi J;, Ramírez M, & Del Pino M. (2022). *AEPCC-Guía: PREVENCIÓN SECUNDARIA DEL CANCER DE CUELLO DEL ÚTERO, 2022. CONDUCTA CLÍNICA ANTE RESULTADOS ANORMALES DE LAS PRUEBAS DE CRIBADO.*
- Torné Blade, A., Saladrigues, M., & Cusido Gimferrer, M. (2014). Guía de cribado del cáncer de cuello de útero en España, 2014. *Revista Española de Patología*, *47*(1), 1-43.
- Vu, T. H., Mousavi, H. S., Monga, V., Rao, G., & Rao, U. A. (2016). Histopathological Image Classification using Discriminative Feature-oriented Dictionary Learning. *IEEE transactions on medical imaging*, *35*(3), 738-751. https://doi.org/10.1109/TMI.2015.2493530
- Vyborny, C. J., & Giger, M. L. (1994). Computer vision and artificial intelligence in

mammography. *American Journal of Roentgenology*, *162*(3), 699-708. https://doi.org/10.2214/ajr.162.3.8109525

- Wang, J., Yu, Y., Tan, Y., Wan, H., Zheng, N., He, Z., Mao, L., Ren, W., Chen, K., Lin, Z., He, G., Chen, Y., Chen, R., Xu, H., Liu, K., Yao, Q., Fu, S., Song, Y., Chen, Q., … Yao, H. (2024). Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer. *Nature Communications*, *15*(1), 4369. https://doi.org/10.1038/s41467-024-48705-3
- Wang, L. (2017). Data Mining, Machine Learning and Big Data Analytics. *International Transaction of Electrical and Computer Engineers System*, *4*(2), 55-61. https://doi.org/10.12691/iteces-4-2-2
- Watanabe, S. (1974). An automated apparatus for cancer prescreening: CYBEST. *Computer Graphics and Image Processing*, *3*(4), 350-358. https://doi.org/10.1016/0146-664X(74)90029-X
- Weinstein, R. S., Graham, A. R., Richter, L. C., Barker, G. P., Krupinski, E. A., Lopez, A. M., Erps, K. A., Bhattacharyya, A. K., Yagi, Y., & Gilbertson, J. R. (2009). Overview of telepathology, virtual microscopy, and whole slide imaging: Prospects for the future. *Human Pathology*, *40*(8), 1057-1069. https://doi.org/10.1016/j.humpath.2009.04.006
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45. https://doi.org/10.1145/365153.365168
- Wied, G. L., Bahr, G. F., Bibbo, M., Puls, J. H., Taylor, J., & Bartels, P. H. (1975). The TICAS-RTCIP real time cell identification processor. *Acta Cytologica*, *19*(3), 286-288.
- Wright, T. C., Stoler, M. H., Behrens, C. M., Sharma, A., Sharma, K., & Apple, R. (2014). Interlaboratory variation in the performance of liquid-based cytology: Insights from the ATHENA trial. *International Journal of Cancer*, *134*(8), 1835-1843. https://doi.org/10.1002/ijc.28514
- Wu, M., Yan, C., Liu, H., & Liu, Q. (2018). Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks. *Bioscience Reports*, *38*(3), BSR20180289. https://doi.org/10.1042/BSR20180289
- Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5987-5995. https://doi.org/10.1109/CVPR.2017.634
- Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., & Yu, S. (2021). A survey on incorporating domain knowledge into deep learning for medical image analysis. *Medical Image Analysis*, *69*, 101985. https://doi.org/10.1016/j.media.2021.101985
- Young, B., Woodford, P., & O'Dowd, G. (2013). *Wheater's Functional Histology Book: A Text and Colour Atlas* (6.ª ed.).
- Younge, P. A. (1965). THE NATURAL HISTORY OF CARCINOMA IN SITU OF THE CERVIX UTERI. *BJOG: An International Journal of Obstetrics and Gynaecology*, *72*(1), 9-12. https://doi.org/10.1111/j.1471-0528.1965.tb01367.x
- Zhang, L., Le Lu, null, Nogues, I., Summers, R. M., Liu, S., & Yao, J. (2017). DeepPap: Deep Convolutional Networks for Cervical Cell Classification. *IEEE Journal of Biomedical and Health Informatics*, *21*(6), 1633-1643. https://doi.org/10.1109/JBHI.2017.2705583
- Zhang, X., Ji, J., Zhang, Q., Zheng, X., Ge, K., Hua, M., Cao, L., & Wang, L. (2025). A large annotated cervical cytology images dataset for AI models to aid cervical cancer screening. *Scientific Data*, *12*(1), 23. https://doi.org/10.1038/s41597-025-04374-5
- Zhang, Y., Ning, C., & Yang, W. (2025). An automatic cervical cell classification model based on improved DenseNet121. *Scientific Reports*, *15*(1), 3240. https://doi.org/10.1038/s41598-025-87953-1

X – ANEXOS

X - ANEXOS

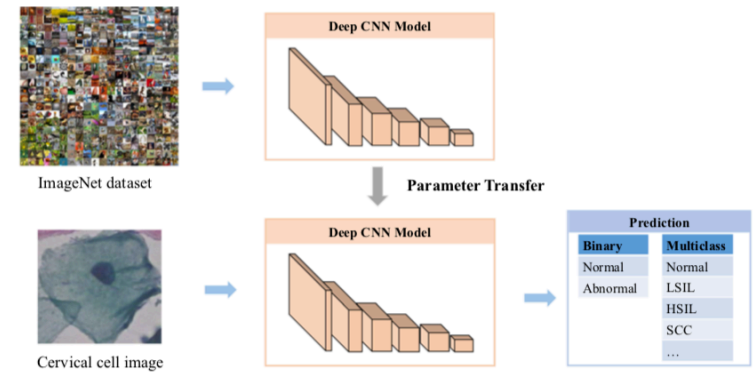
El entrenamiento de las redes CNN se puede realizar de 3 formas distintas que han sido desarrolladas previamente en este trabajo en el apartado de *introducción*. En el **Anexo 1** (Jiang et al., 2023), se expone de manera esquemática 3 ejemplos de estos métodos de aprendizaje aplicados al diagnóstico de citologías cervicovaginales:

1. Aprendizaje por transferencia (**Anexo 1-a**). Este ejemplo, se emplea la base de datos de imágenes de ImageNet (que contiene numerosas imágenes de la naturaleza y de objetos) para entrenar a la red CNN. Aprovechando que esta red preentrenada es capaz de identificar y etiquetar imágenes de la naturaleza, se aplica una *transferencia de parámetros* para que la nueva red CNN sea capaz de detectar características generales de la imagen como colores, formas y texturas. Tras un ajuste de parámetros y un nuevo entrenamiento con imágenes de células, la nueva red CNN podrá hacer una predicción sobre la imagen celular tanto binaria como multiclase.

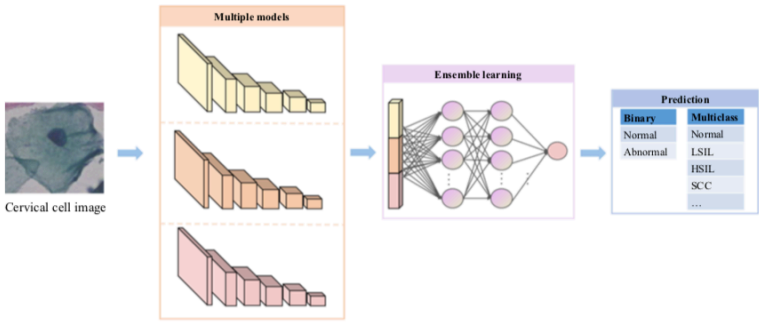
2. Conjunto multimodelo (**Anexo 1-b**). Se emplean redes CNN independientes con un entrenamiento previo independiente de cada una de ellas, que se combinan en un *aprendizaje conjunto* para emitir en la capa de salida una predicción final tanto binaria como multiclase.

3. Fusión de características híbridas (**Anexo 1-c**). En este ejemplo, se emplea una célula cervical que sigue dos vías simultáneas. Por un lado accede a la red CNN y por otro accede a una etapa de segmentación en el que se le aplica una máscara celular previa a su extracción manual de características. Posteriormente se produce una extracción de características para la elaboración de una predicción final tanto binaria como multiclase.

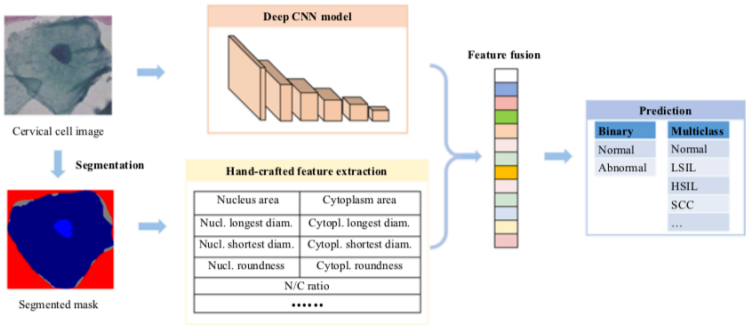
Anexo 2. Enfoques de entrenamiento de CNN. Imagen obtenida de la revisión sistemática: "A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis" de Jiang y cols. (Jiang et al., 2023).



(a) Transfer learning based identification



(b) Multi-model ensemble based identification



(c) Multi-type feature fusion based identification

En el **Anexo 2** se detallan las fases de implantación de un diagnóstico automatizado en los laboratorios de Anatomía Patológica, que proponen Jiang y cols. (Jiang et al., 2023). El diagrama se desarrolla desde la parte inferior de la figura hasta la superior, dividido en 3 fases:

Primera parte: adquisición y preprocesamiento de la imagen.

- Recogida de imágenes de alta calidad, mediante un microscopio óptico, imágenes digitales o un escaneado automático del portaobjetos mediante escáneres especializados. Una vez se obtiene la imagen digital completa del portaobjetos, se divide en celdillas uniformes para facilitar el trabajo del algoritmo, que pasará a trabajar con imágenes de la preparación digital.

- Posteriormente se realiza un preprocesamiento de las imágenes que consiste en una mejora de la calidad de la imagen en la que se trata de eliminar cualquier artefacto o ruido que pueda alterar su interpretabilidad. Se consigue mediante la reducción de ruido, normalización de las manchas y mejora del contraste. Nuestro estudio no sigue esta fase de preprocesamiento tal y como se ha expuesto anteriormente en este trabajo.

- Detección de áreas de interés mediante el empleo de algoritmos para detectar y clasificar áreas de interés para su posterior análisis (celdillas con imágenes de células o áreas sospechosas).

- Segmentación celular. Cuando se han detectado estas imágenes con áreas de interés se pueden segmentar en células individuales para intentar identificar las regiones que contienen las células sospechosas. Este paso no es imprescindible, pero presenta grandes ventajas, como aportar mayor detalle y precisión.

Segunda parte: análisis de la imagen.

- Extracción de características. Se extrae de cada célula individual sus características biomédicas claves para el diagnóstico; como el tamaño del núcleo, color, textura, forma, perímetro, posición del mismo en el citoplasma y área que representa el núcleo respecto al citoplasma (relación núcleo-citoplasma).

- Identificación de células. Con toda esta información, se clasifican las células en las distintas categorías diagnósticas del Sistema Bethesda, mediante árboles de decisión, máquinas de vectores de soporte, o modelos DL.

- Diagnóstico de todo el portaobjetos. Cuando se ha obtenido la clasificación a nivel celular (de célula o parche segmentado), todos los datos se integran para realizar una predicción a nivel del portaobjetos. Se puede hacer de dos formas: fusionando las características de cada célula para generar una característica a nivel de portaobjetos, o combinando todas las probabilidades de predicción para generar la probabilidad final de diagnóstico a nivel del portaobjetos.

Tercera parte: elaboración de un informe.

- Revisión y validación de los resultados por un patólogo experto.

- Informe con los resultados. Aporta los resultados observados tras analizar todo el portaobjetos, que puede incluir la ubicación o características de las células alteradas o el nivel de sospecha del resultado observado en el portaobjetos.

Anexo 5. Fases de implantación del diagnóstico automatizado en citología cervicovaginal. Imagen obtenida de la revisión sistemática: "A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis" de Jiang y cols. (Jiang et al., 2023).